

**This Page Is Inserted by IFW Operations  
and is not a part of the Official Record**

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problems Mailbox.**

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 30, 2002, 08:15:08 ; Search time 68.88 Seconds  
(without alignments)  
64,941 Million cell updates/sec

Title: US-09-095-478A-8

Perfect score: 673

Sequence: 1 DFWGMWNNCNVAMITRE.....VRKSHITGPIVHCTAGVR 122

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	268.5	39.9	2485	1 PTND_HUMAN	Q12923 homo sapien
2	266.5	39.6	377	1 PTP2_DICDI	P24138 dictyosteli
3	250	37.1	1	1 PTN3_HUMAN	P26045 homo sapien
4	244	36.3	2316	1 PTP2_RAT	O62656 rattus norv
5	239	35.5	1337	1 PTPJ_HUMAN	Q12913 homo sapien
6	238	35.4	2314	1 PTPJ_HUMAN	P23471 homo sapien
7	236.5	35.1	1452	1 PTP2_HUMAN	P28827 homo sapien
8	235.5	35.0	1152	1 CD45_MOUSE	P06800 mus musculu
9	235.5	35.0	1452	1 PTPM_MOUSE	P28828 mus musculu
10	233.5	34.7	1255	1 CD45_RAT	P04157 rattus norv
11	232	34.5	1187	1 PTNE_HUMAN	Q15678 homo sapien
12	232	34.5	1238	1 PTPJ_MOUSE	Q64455 mus musculu
13	232	34.5	1912	1 PTPD_HUMAN	P23468 homo sapien
14	231	34.3	1997	1 PTPB_HUMAN	P23467 homo sapien
15	230.5	34.2	1304	1 CD45_HUMAN	P08575 homo sapien
16	230.5	34.2	1430	1 PTPU_HUMAN	Q92729 homo sapien
17	230	34.2	1897	1 PTPF_HUMAN	P10586 homo sapien
18	229.5	34.1	382	1 PTP2_MOUSE	Q06180 mus musculu
19	229	34.0	2029	1 LAR_DROME	P16621 drosophila
20	228.5	34.0	1026	1 PTP1_CAEEL	P28191 caenorhabdi
21	225.5	33.5	363	1 PTN2_RAT	P35233 rattus norv
22	224.5	33.4	415	1 PTN2_HUMAN	P17706 homo sapien
23	224	33.3	775	1 PTNC_MOUSE	P35831 mus musculu
24	223.5	33.2	1176	1 PTNL_MOUSE	Q62136 mus musculu
25	222	33.0	1189	1 PTNE_MOUSE	Q62130 mus musculu
26	221.5	32.9	1705	1 PTPO_MOUSE	P70289 mus musculu
27	221	32.8	780	1 PTNC_HUMAN	O05209 homo sapien
28	220.5	32.8	1439	1 PTPK_HUMAN	Q15262 homo sapien
29	220.5	32.8	1457	1 PTPK_MOUSE	P35822 mus musculu
30	219	32.5	845	1 CSW_DROME	P29349 drosophila
31	218	32.4	434	1 PTN1_CHICK	O13016 gallus gall
32	218	32.4	435	1 PTN1_HUMAN	P18031 homo sapien
33	217.5	32.3	1174	1 PTNL_HUMAN	Q15825 homo sapien

34	217.5	32.3	1175	1 PTNL_RAT	Q62728 rattus norv
35	217.5	32.3	1445	1 PTPG_HUMAN	P23470 homo sapien
36	216	32.1	432	1 PTN1_RAT	P20417 rattus norv
37	215.5	32.0	1711	1 PTPO_RAT	Q64612 rattus norv
38	214.5	31.9	926	1 PTN4_HUMAN	P29074 homo sapien
39	214	31.8	432	1 PTN1_MOUSE	P35821 mus musculu
40	210.5	31.3	593	1 PTN9_HUMAN	P43378 homo sapien
41	207.5	30.8	979	1 PTPN_HUMAN	Q16849 homo sapien
42	207	30.8	360	1 PTN7_MOUSE	P32336 homo sapien
43	206	30.6	595	1 PTN6_MOUSE	P29351 mus musculu
44	205.5	30.5	979	1 PTPN_BOVIN	P56722 bos taurus
45	205.5	30.5	979	1 PTPN_MOUSE	Q60673 mus musculu

ALIGNMENTS

RESULT 1

ID	PTND_HUMAN	STANDARD;	PRT; 2485 AA.
AC	Q12923; Q15263; Q16826; Q15264; Q15265; Q15159;		
DT	20-AUG-2001 (Rel. 40, Created)		
DT	20-AUG-2001 (Rel. 40, Last sequence update)		
DT	20-AUG-2001 (Rel. 40, Last annotation update)		
DE	PROTEIN TYROSINE PHOSPHATASE, NON-RECEPTOR TYPE 13 (EC 3.1.3.48)		
DE	(PROTEIN-TYROSINE PHOSPHATASE 1E) (PTP-EI) (PTP-BAS) (PROTEIN-TYROSINE PHOSPHATASE PTP1)		
DE	(FAP-1).		
GN	PTPN13 OR PTP1E OR PTP1L OR PNP1.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID:9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE-Breast carcinoma;		
RX	MEDLINE-94350988; PubMed-8071359;		
RA	Bayville D., Ahmad S., Stocco R., Shen S.-H.;		
RT	"A novel protein-tyrosine phosphatase with homology to both the cytoskeletal proteins of the band 4.1 family and junction-associated		
RT	quanylate kinases."		
RL	J. Biol. Chem. 269:22320-22327(1994).		
RP	SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.		
RN	[2]		
RP	TISSUE-Leukemia;		
RC	MEDLINE-94116679; PubMed-8287977;		
RX	Maekawa K., Imagawa N., Nagamatsu M., Harada S.;		
RA	"Molecular cloning of a novel protein-tyrosine phosphatase containing		
RT	a membrane-binding domain and GLGF repeats."		
RT	FERS Lett. 337:200-206(1994).		
RN	[3]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE-Fibroblast;		
RX	MEDLINE-95014139; PubMed-7929060;		
RA	Satas J., Claesson-Welsh L., Heldin C.-H., Gonen L.J.;		
RT	"Cloning and characterization of PTP1, a protein tyrosine phosphatase		
RT	with similarities to cytoskeletal-associated proteins."		
RL	J. Biol. Chem. 269:24082-24089(1994).		
RN	[4]		
RP	SEQUENCE OF 1216-2490 FROM N.A.		
RC	TISSUE-Pancreas;		
RA	Wang H.Y.;		
RL	Submitted (JUN-1994) to the EMBL/GenBank/DBJ databases.		
RN	[5]		
RP	STRUCTURE BY NMR OF 1361-1456.		
RX	MEDLINE-20170882; PubMed-10704206;		
RA	Kozlov G., Gehring K., Ekiel I.;		
RT	"Solution structure of the Pdz2 domain from human phosphatase hPTP1E		
RT	and its interactions with C-terminal peptides from the Fas		
RT	receptor."		
RL	Biochemistry 39:2572-2580(2000).		
CG	FUNCTION: BINDS TO A NEGATIVE REGULATORY DOMAIN IN FAS THAT		
CC	INHIBITS FAS-INDUCED APOPTOSIS.		





```

KW Hydrolase.
FT DOMAIN 78 86 POLY-ASP.
FT DOMAIN 95 103 POLY-ASN.
FT ACT_SITE 281 BY SIMILARITY.
SQ SEQUENCE 377 AA; 43488 MW; 5B08F6EB54829FD9 CRC64;

Query Match 39.6%; Score 266.5; DB 1; Length 377;
Best Local Similarity 40.7%; Pred. No. 9.5e-22;
Matches 50; Conservative 20; Mismatches 50; Indels 3; Gaps

QY 1 DFWGMNENNCNVAMITREIEGGVIKCCSYMPVSLKEPFEKFKHFFHLLLE-NFOITQYFV 59
DB 167 DFWKMIWEQNSNIIVMLTREENFKTKCDKXWPD--KDEERYGNFVFKDFNNITPIDILI 224
QY 60 IRIFOIVKKSCTGSKSVKHLQFIKWPDHGTSPASVDFFIKYVRYVRKSHITGPLVLVHCTAG 119
DB 225 RREFTEFLNLKDNKTRKVIHFQYTTWPDHGTVPSTGLFKFVSFVDHKEKSGPIVVHCSAG 284
QY 120 VGR 122
DB 285 IGR 287

RESULT 3
PTN3_HUMAN
ID PTN3_HUMAN STANDARD; PRT; 913 AA.
AC P26045;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROTEIN TYROSINE PHOSPHATASE, NON-RECEPTOR TYPE 3 (EC 3.1.3.48)
DE (PROTEIN-TYROSINE PHOSPHATASE H1) (PTP-H1).
DE PTPN3 OR PTPH1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxId=9606;
RN [1]
RX MEDLINE=91296738; PubMed=1648725;
RA Yang Q., Tonks N.K.;
RT "Isolation of a cDNA clone encoding a human protein-tyrosine
RT phosphatase with homology to the cytoskeletal-associated proteins
RT band 4.1, ezrin, and talin."
RL Proc. Natl. Acad. Sci. U.S.A. 88:5949-5953(1991).
RN [2]
RX SEQUENCE OF 194-896 FROM N.A.
RN TISSUE-Colon;
RX MEDLINE=92327504; PubMed=1626183;
RA Arimura Y., Hinoda Y., Itoh F., Takekawa M., Tsujisaki M., Adachi M.,
RA Imai K., Yachi A.;
RT "cDNA cloning of new protein tyrosine phosphatases in the human
RT colon."
RL Tumour Biol. 13:180-186(1992).
RN [3]
RX SEQUENCE OF 899-913 FROM N.A.
RX MEDLINE=95179278; PubMed=7874267;
RA Ikuta S., Itoh F., Hinoda Y., Toyota M., Makiguchi Y., Imai K.,
RA Yachi A.;
RT "Expression of cytoskeletal-associated protein tyrosine phosphatase
RT PTPH1 mRNA in human hepatocellular carcinoma."
RL J. Gastroenterol. 29:727-732(1994).
CC -1- FUNCTION: MAY ACT AT JUNCTIONS BETWEEN THE MEMBRANE AND THE
CC CYTOSKELETON
CC -1- CATALYTIC ACTIVITY: PROTEIN TYROSINE PHOSPHATE + H(2)O =
CC PROTEIN TYROSINE + ORTHOPHOSPHATE.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- SIMILARITY: BELONGS TO THE NON-RECEPTOR CLASS OF THE PROTEIN-
CC TYROSINE PHOSPHATASE FAMILY.
CC -1- SIMILARITY: CONTAINS 1 BAND 4.1-LIKE DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.

```

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

EMBL; M64572; AAA35647.1; -  
EMBL; S39392; AAB22439.2; -  
EMBL; S76309; AAB33583.1; -  
PIR; A41109; A41109.  
HSP; P18031; IPTT.  
MIL; 176877; -  
InterPro: IPR000299; Band\_4.1.  
InterPro: IPR001478; PDZ.  
InterPro: IPR000387; TYR\_phosphatase.  
InterPro: IPR000242; Tyr\_prot\_phphatase.  
Pfam; PF00373; Band\_41; 1.  
Pfam; PF00595; PDZ; 1.  
Pfam; PF00102; Y\_phosphatase; 1.  
PRINTS; PR00700; PRTPHPHTASE.  
PRINTS; PR00935; BAND41.  
SMART; SM00295; B41; 1.  
SMART; SM00228; PDZ; 1.  
SMART; SM00194; PTPc; 1.  
PROSITE; PS00660; BAND\_41\_1; 1.  
PROSITE; PS00661; BAND\_41\_2; 1.  
PROSITE; PS50057; BAND\_41\_3; 1.  
PROSITE; PS50106; PDZ; 1.  
PROSITE; PS00383; TYR\_PHOSPHATASE\_1; 1.  
PROSITE; PS50056; TYR\_PHOSPHATASE\_2; 1.  
PROSITE; PS50055; TYR\_PHOSPHATASE\_PTP; 1.  
Structural protein; Cytoskeleton; Hydrolase.  
FT DOMAIN 84 241 BAND 4.1-LIKE.  
FT DOMAIN 510 582 PDZ.  
FT DOMAIN 670 913 PROTEIN-TYROSINE PHOSPHATASE.  
FT ACT\_SITE 842 842 BY SIMILARITY.  
SQ SEQUENCE 913 AA; 104029 NW; 29A539ACDE2F1515 CRC64;

Query Match 37.1%; Score 250; DB 1; Length 913;  
Best Local Similarity 37.9%; Pred. No. 1.6e-19;  
Matches 47; Conservative 23; Mismatches 48; Indels 6; Gaps 3;

QY 2 FWMGWNENCNVAMITREIEGGVICKCSWPVSLKPELEKH--RHVLLENFQITQYFV 59  
DB 728 FQVYVWDQKLSLVMLVTLTERGKTCQCHQWP--DPDVMNHGGFHQCSDECTIAYV 784  
QY 60 IRIFQIVKSTGKSHSVKHLQFKWPDHGPASVDFEIKYRVVRKSHI-TGPLLVHCTA 118  
DB 785 SREMLVNTQTGEHTVTHLQYVAPWDHGPDDOSSDFEFNVYRSLRVDSPEVLVHCSA 844  
QY 119 GVGR 122  
DB 845 GIGR 848

RESULT : 4  
PTPZ\_RAT  
ID PTPZ\_RAT STANDARD; PRT: 2316 AA.  
AC Q62656; Q62621;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE PROTEIN-TYROSINE PHOSPHATASE ZETA PRECURSOR (EC 3.1.3.48) (R-PTP-  
DE ZETA) (PHOSPHACAN) (3F8 CHONDROITIN SULFATE PROTEOGLYCAN) (3H1 KERATAN  
DE SULFATE PROTEOGLYCAN).  
GN PTPZ1 OR PTPZ OR PTPZ.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.  
OC NCBI\_TaxId=10116;



AC Q12913;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DE 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE PROTEIN-TYROSINE PHOSPHATASE ETA PRECURSOR (EC 3.1.3.48) (R-PTP-ETA)  
 DE (DENSITY ENHANCED PHOSPHATASE-1) (DEP-1) (CD148 ANTIGEN).  
 GN PTPRJ OR DEPI  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP MEDLINE=95024024; PubMed=7937872;  
 RA Cesman A., Yang Q., Tonks N.K.;  
 RT "Expression of DEP-1, a receptor-like protein-tyrosine-phosphatase,  
 is enhanced with increasing cell density.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 91:9680-9684(1994).  
 CC !- FUNCTION: MAY CONTRIBUTE TO THE MECHANISM OF CONTACT INHIBITION OF  
 CELL GROWTH.  
 CC !- CATALYTIC ACTIVITY: PROTEIN TYROSINE PHOSPHATE + H(2)O =  
 PROTEIN TYROSINE + ORTHOPHOSPHATE.  
 CC !- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
 CC !- PTM: N- AND O-GLYCOSYLATED.  
 CC !- SIMILARITY: CONTAINS 1 PROTEIN-TYROSINE PHOSPHATASE DOMAIN.  
 CC !- SIMILARITY: CONTAINS 5 FIBRONECTIN TYPE III-LIKE DOMAINS.  
 CC !- DATABASE: NAME=PROW; NOTE=CD guide CD148 entry;  
 CC WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cd148.htm".  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: U10886; AAB36687.1;  
 DR HSSP: P18052; 1YFO.  
 DR MIM: 600925;  
 DR InterPro: IPR001777; FN-III.  
 DR InterPro: IPR000387; TYR\_phosphatase.  
 DR InterPro: IPR000242; Tyr\_prot\_phptase.  
 DR Pfam: PF00041; fn3; 5.  
 DR Pfam: PF00102; Y-phosphatase; 1.  
 DR PRINTS: PRO0700; PRTYPHPTASE.  
 DR SMART: SM00060; FN3; 6.  
 DR SMART: SM00194; PTPC; 1.  
 DR PROSITE: PS00383; TYR\_PHOSPHATASE\_1; 1.  
 DR PROSITE: PS00556; TYR\_PHOSPHATASE\_2; 1.  
 DR PROSITE: PS50055; TYR\_PHOSPHATASE\_PTP; 1.  
 KW Signal; Glycoprotein; Transmembrane; Repeat; Hydrolase.  
 FT SIGNAL 1 35  
 FT CHAIN 36 1337  
 FT DOMAIN 36 975  
 FT TRANSMEM 976 996  
 FT DOMAIN 997 1337  
 FT DOMAIN 119 199  
 FT DOMAIN 366 446  
 FT DOMAIN 454 532  
 FT DOMAIN 540 615  
 FT DOMAIN 626 710  
 FT DOMAIN 1065 1337  
 FT ACT\_SITE 1239 1339  
 FT CARBOHYD 72 72  
 FT CARBOHYD 82 82  
 FT CARBOHYD 93 93  
 FT CARBOHYD 104 104  
 FT CARBOHYD 142 142  
 FT CARBOHYD 172 172  
 FT CARBOHYD 192 192  
 FT CARBOHYD 231 231  
 FT CARBOHYD 258 258

ET CARBOHYD 278 278 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 ET CARBOHYD 342 342 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 ET CARBOHYD 351 351 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 ET CARBOHYD 376 376 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 ET CARBOHYD 391 391 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 ET CARBOHYD 396 396 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 ET CARBOHYD 413 413 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 ET CARBOHYD 431 431 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 ET CARBOHYD 501 501 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 ET CARBOHYD 525 525 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 ET CARBOHYD 536 536 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 ET CARBOHYD 582 582 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 ET CARBOHYD 603 603 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 ET CARBOHYD 618 618 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 ET CARBOHYD 628 628 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 ET CARBOHYD 637 637 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 ET CARBOHYD 666 666 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 ET CARBOHYD 669 669 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 ET CARBOHYD 761 761 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 ET CARBOHYD 772 772 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 ET CARBOHYD 784 784 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 ET CARBOHYD 790 790 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 ET CARBOHYD 824 824 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 ET CARBOHYD 910 910 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 ET CARBOHYD 937 937 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 1337 AA; 145985 MW; 5686DE6D1F64236E CRC64;  
 Query Match 35.5%; Score 239; DB 1; Length 1337;  
 Best Local Similarity 38.1%; Pred. No. 3 8e-18;  
 Matches 48; Conservative 22; Mismatches 50; Indels 6; Gaps 3;  
 QY 1 DFWGMWNNVNCVIAITREIEGGVTKCSYMPVSLKEPFLKHFHVLLENFOITQYFVI 60  
 DB 1122 DFWRMVWZNVVIAIMLTKEVQGRTKCEYWP--SKQAQDYGDITVAMTSEIVLPEWTI 1179  
 QY 61 RIFQIVKKSSTGSHSVKHLQFTKWPDPGTPASVDFFIKY---VR-VYRSHITGPIILVHC 116  
 DB 1190 RDTVANIOTSESHPLRQFHFTSWPDHGVDPDTDLLINRYLVRYDMKOSPPSPILVHC 1239  
 QY 117 TAGVGR 122  
 DB 1240 SAGVGR 1245  
 RESULT: 6  
 PTPZ\_HUMAN  
 ID PTPZ\_HUMAN STANDARD; PRT: 2314 AA.  
 AC P23471;  
 DT 01-NOV-1991 (Rel. 20, Created)  
 DT 01-JUL-1993 (Rel. 26, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE PROTEIN-TYROSINE PHOSPHATASE ZETA PRECURSOR (EC 3.1.3.48) (R-PTP-  
 ZETA).  
 GN PTPZ1 OR PTPRZ OR PTPZ.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [14]  
 RP SEQUENCE FROM N.A. (LONG FORM).  
 RC TISSUE=Brain;  
 RX MEDLINE=92365472; PubMed=1323835;  
 RA Krueger N.X., Saito H.;  
 RT "A human transmembrane protein-tyrosine-phosphatase, PTP zeta, is  
 expressed in brain and has an N-terminal receptor domain homologous  
 to carbonic anhydrases.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 89:7417-7421(1992).  
 RN [2]  
 RP SEQUENCE FROM N.A. (LONG AND SHORT FORMS).  
 RC TISSUE=Brain stem;  
 RX MEDLINE=93252948; PubMed=8387522;  
 RA Levy J.B., Canoll P.D., Silvennoinen O., Barnea G., Morse B.,

RA Honegger A.M., Huang J.-T., Cannizzaro L.A., Park S.-H., Druck T.,  
 RT Huebner K., Sap J., Ehrlich M., Musacchio J.M., Schlessinger J.;  
 RA "The cloning of a receptor-type protein tyrosine phosphatase  
 RT expressed in the central nervous system.";  
 J. Biol. Chem. 268:10573-10581(1993).  
 RN [3]  
 RP SEQUENCE OF 1479-2091 FROM N.A. (LONG FORM).  
 RC TISSUE=Liver;  
 RX MEDLINE=91006018; PubMed=2170109;  
 RA Krueger N.X., Streuli M., Saito H.;  
 RT "Structural diversity and evolution of human receptor-like protein  
 RT tyrosine phosphatases";  
 J. Biol. Chem. 268:10573-10581(1993).  
 RN [4]  
 RP SEQUENCE OF 1749-1990 AND 2047-2280 FROM N.A. (LONG FORM).  
 RC TISSUE=Brain stem;  
 RX MEDLINE=90384936; PubMed=2169617;  
 RA Kaplan R., Morse B., Huebner K., Croce C., Hawk R., Ravera M.,  
 RA Ricca G., Jaye M., Schlessinger J.;  
 RT "Cloning of three human tyrosine phosphatases reveals a multigene  
 RT family of receptor-linked protein-tyrosine-phosphatases expressed in  
 RT brain.";  
 Proc. Natl. Acad. Sci. U.S.A. 87:7000-7004(1990).  
 CC -1- FUNCTION: MAY BE INVOLVED IN THE REGULATION OF SPECIFIC  
 CC DEVELOPMENTAL PROCESSES IN THE CNS.  
 CC -1- CATALYTIC ACTIVITY: PROTEIN TYROSINE PHOSPHATE + H(2)O =  
 CC PROTEIN TYROSINE + ORTHOPHOSPHATE.  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: A LONG FORM (SHOWN HERE) AND A  
 CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.  
 CC -1- TISSUE SPECIFICITY: SPECIFICALLY EXPRESSED IN THE CENTRAL NERVOUS  
 CC SYSTEM, WHERE IT IS LOCALIZED IN THE PURKINJE CELL LAYER OF THE  
 CC CEREBELLUM, THE DENTATE GYRUS, AND THE SUBEPENDYMAL LAYER OF THE  
 CC ANTERIOR HORN OF THE LATERAL VENTRICLE. DEVELOPMENTALLY REGULATED  
 CC IN THE BRAIN.  
 CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE EUKARYOTIC-  
 CC TYPE CARBONIC ANHYDRASE FAMILY.  
 CC -1- SIMILARITY: CONTAINS 2 PROTEIN-TYROSINE PHOSPHATASE DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 1 FIBRONECTIN TYPE III-LIKE DOMAIN.  
 CC -1- CAUTION: CALLED RTTPASE BETA IN REF.2 AND 4.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M93426; AAA60225.1; -;  
 DR EMBL; X54135; CAA38070.1; -;  
 DR PIR; S12054; S12054.  
 DR PIR; A46151; A46151.  
 DR HSP; P18052; IYFO.  
 DR MIM; 176891; -;  
 DR InterPro; IPR001148; Carb\_anhydrase.  
 DR InterPro; IPR001777; FN\_III.  
 DR InterPro; IPR000387; Tyr\_phosphatase.  
 DR InterPro; IPR000242; Tyr\_prot\_phphatase.  
 DR Pfam; PF00194; carb\_anhydrase; 1.  
 DR Pfam; PF00041; fn3; 1.  
 DR Pfam; PF00102; Y\_phosphatase; 2.  
 DR PRINTS; PR00700; PRTYPHPTASE.  
 DR ProDom; PD000865; Carb\_anhydrase; 1.  
 DR SMART; SM00060; FN3; 1.  
 DR SMART; SM00194; PTPC; 2.  
 DR PROSITE; PS00383; Tyr\_PHOSPHATASE\_1; 1.  
 DR PROSITE; PS50056; Tyr\_PHOSPHATASE\_2; 2.  
 DR PROSITE; PS50035; Tyr\_PHOSPHATASE\_PTP; 2.  
 KW Signal; Glycoprotein; Transmembrane; Hydrolase; Alternative splicing.  
 FT SIGNAL 1 24 BY SIMILARITY.  
 FT CHAIN 25 2314 PROTEIN-TYROSINE PHOSPHATASE ZETA.  
 FT DOMAIN 25 1635 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 1636 1661 POTENTIAL.  
 FT DOMAIN 1662 2314 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 34 302 CARBONIC-ANHYDRASE LIKE.  
 FT DOMAIN 312 406 FIBRONECTIN TYPE-III.  
 FT DOMAIN 1744 1997 PROTEIN-TYROSINE PHOSPHATASE 1.  
 FT DOMAIN 1998 2314 PROTEIN-TYROSINE PHOSPHATASE 2.  
 FT BINDING 587 587 CHONDROITIN SULFATE (POTENTIAL).  
 FT BINDING 637 637 CHONDROITIN SULFATE (POTENTIAL).  
 FT BINDING 997 997 CHONDROITIN SULFATE (POTENTIAL).  
 FT BINDING 1548 1548 CHONDROITIN SULFATE (POTENTIAL).  
 FT BINDING 1550 1550 CHONDROITIN SULFATE (POTENTIAL).  
 FT ACT\_SITE 1932 1932 CHONDROITIN SULFATE (POTENTIAL).  
 FT SITE -2222 2222 BY SIMILARITY.  
 FT CARBOHYD 105 105 ANCESTRAL ACTIVE SITE.  
 FT CARBOHYD 134 134 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 223 223 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 232 232 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 324 324 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 381 381 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 497 497 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 501 501 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 552 552 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 602 602 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 629 629 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 677 677 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1017 1017 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1050 1050 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1082 1082 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1122 1122 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1456 1456 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1561 1561 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 1617 1617 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT VARSPLIC 755 1614 MISSING (IN SHORT ISOFORM).  
 FT CONFLICT 1722 1728 MISSING (IN REF. 2).  
 SQ SEQUENCE 2314 AA; 254528 MW; 77DBDEF4A0F5FB42 CRC64;

Query Match 35.4%; Score 238; DB 1; Length 2314;  
 Best Local Similarity 37.0%; Pred. No. 9e-18;  
 Matches 50; Conservative 24; Mismatches 43; Indels 18; Gaps 5;

QY 1 DFVGMWENCNVAMITREIEGVVKCCSWPVSLEKPELFKHVHLENFOITQYFVI 60  
 DB 1809 DFRMIWEHNEVILVMTNLVKGRRKCDQIWPADGSE--EYGNFLVTQKSVQLAYTV 1866  
 QY 61 RIFQI-----VKKSTGKSHS-----VKHLQFIKWPDHGPPASVDFFIKYVRVRKS-----H 107  
 DB 1867 RNFTLRNTRIKKGSQKGRPSRVVTVQHYTQWPDGVP---EYSLPVLTVFVRKAAAYAKRH 1923

QY 108 ITGPLELHVCTAGVGR 122  
 DB 1924 AVGPVVVHCAGVGR 1938

RESULT 7  
 PTPM\_HUMAN STANDARD; PRT; 1452 AA.  
 AC P28827;  
 DT 01-DEC-1992 (Rel. 24, Created)  
 DT 01-DEC-1992 (Rel. 24, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE PROTEIN-TYROSINE PHOSPHATASE MU PRECURSOR (EC 3.1.3.48) (R-PTP-MU).  
 GN PTPRM OR PTPRL1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92008644; PubMed=1655529;  
 RA Gebbink M.F.B.G., van Etten I., Hateboer G., Suijkerbuijk R.,  
 RA Beijersbergen R., Geurts van Kessel A., Moolenaar W.H.;  
 RT "Cloning, expression and chromosomal localization of a new putative

```

receptor-like protein tyrosine phosphatase." ;
FEBS Lett. 290:123-130(1991).
(2)
X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS) OF 879-1156.
MEDLINE-98010572; PubMed-9346878;
Hoffmann K.M., Tonks N.K., Barford D.;
"The crystal structure of domain 1 of receptor protein-tyrosine
phosphatase mu." ;
J. Biol. Chem. 272:27505-27508(1997).
CC -!- FUNCTION: MAY PLAY A KEY ROLE IN SIGNAL TRANSDUCTION AND GROWTH
CC CONTROL.
CC -!- CATALYTIC ACTIVITY: PROTEIN TYROSINE PHOSPHATE + H2O =
CC PROTEIN TYROSINE + ORTHOPHOSPHATE.
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -!- SIMILARITY: CONTAINS 1 MAM DOMAIN.
CC -!- SIMILARITY: CONTAINS 4 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS 4 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS 2 PROTEIN-TYROSINE PHOSPHATASE DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X58288; CAA41226.1; -
CC PIR; S17669; S17669.
CC PDB; 1RPM; 01-APR-98.
CC MIM; 176886; -
CC InterPro: IPR001777; FN_III.
CC InterPro: IPR003599; Ig.
CC InterPro: IPR003006; Ig_MHC.
CC InterPro: IPR000998; MAM.
CC InterPro: IPR000387; TYR_phosphatase.
CC InterPro: IPR000242; Tyr_prot_phphatase.
CC Pfam; PF00041; fn3; 3.
CC Pfam; PF00047; ig; 1.
CC Pfam; PF00629; MAM; 1.
CC Pfam; PF00102; Y_phosphatase; 2.
CC PRINTS; PR00014; FNYPEIII.
CC PRINTS; PR00020; MAMDOMAIN.
CC PRINTS; PR00700; PRTPHPHTASE.
CC SMART; SM00060; FN3; 2.
CC SMART; SM00409; IG; 1.
CC SMART; SM00137; MAM; 1.
CC SMART; SM00194; PTPC; 2.
CC PROSITE; PS00383; TYR_PHOSPHATASE_1; 2.
CC PROSITE; PS00506; TYR_PHOSPHATASE_2; 2.
CC PROSITE; PS00555; TYR_PHOSPHATASE_PTP; 2.
CC PROSITE; PS00740; MAM_1; 1.
CC PROSITE; PS00600; MAM_2; 1.
CC Hydrolase; Receptor; Glycoprotein; Signal; Transmembrane;
KW Immunoglobulin domain; Repeat; 3D-structure.
KW SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 1452 PROTEIN-TYROSINE PHOSPHATASE MU.
FT DOMAIN 21 742 EXTRACELLULAR (POTENTIAL).
FT TRANSEM 743 764 POTENTIAL.
FT DOMAIN 765 1452 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 22 184 MAM.
FT DOMAIN 199 267 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 287 374 FIBRONECTIN TYPE-III 1.
FT DOMAIN 383 466 FIBRONECTIN TYPE-III 2.
FT DOMAIN 486 571 FIBRONECTIN TYPE-III 3.
FT DOMAIN 589 671 FIBRONECTIN TYPE-III 4.
FT DOMAIN 923 1153 PROTEIN-TYROSINE PHOSPHATASE 1.
FT DOMAIN 1213 1447 PROTEIN-TYROSINE PHOSPHATASE 2.
FT ACT_SITE 1095 1095 BY SIMILARITY.
FT ACT_SITE 1389 1389 BY SIMILARITY.
FT DISULFID 206 260 POTENTIAL.
FT CARBOHYD 72 72 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 92 92 N-LINKED (GLCNAC. . .) (POTENTIAL).

```

CARBOHYD 131 131 N-LINKED (GLCNAC. .) (POTENTIAL).  
 249 249 N-LINKED (GLCNAC. .) (POTENTIAL).  
 406 406 N-LINKED (GLCNAC. .) (POTENTIAL).  
 414 414 N-LINKED (GLCNAC. .) (POTENTIAL).  
 454 454 N-LINKED (GLCNAC. .) (POTENTIAL).  
 534 534 N-LINKED (GLCNAC. .) (POTENTIAL).  
 544 544 N-LINKED (GLCNAC. .) (POTENTIAL).  
 598 598 N-LINKED (GLCNAC. .) (POTENTIAL).  
 651 651 N-LINKED (GLCNAC. .) (POTENTIAL).  
 681 681 N-LINKED (GLCNAC. .) (POTENTIAL).  
 SEQUENCE 1452 AA; 163633 MW; 0817503595A6C7E0 CRC64;

Query: Watch 35.1%; Score 236.5; DB 1; Length 1452;  
 Best local Similarity 36.3%; Pred. No. 7.8e-18;  
 Matches 45; Conservative 19; Mismatches 55; Indels 5; Gaps

QY 161 DFWGMWNNCNVIAITRETEGGVIKCCSWPVSLEKPEFKHFHVLLENFQITQYFVI 60  
 DB 981 DFWRMVHENTASIIIMVNLVEGVKCKCKYMP---DDTEYKDKIVLTETELLAIEYVI 1037  
 QY 61 RIFQIVKKSSTGSHSVKHLQFTKWDHGTDPASVDVFFIKYVRYVRKSH--ITGPLLVHCTA 118  
 DB 1088 RTFAVEKRGVHEIRFQHFHTGWDHGVPHYATYGLLGFVRQVKSKSPPSAGPLVWHCSA 1097  
 QY 119 GVGVR 122  
 DB 1098 GAGR 1101

RESULT 8  
 CD45\_MOUSE STANDARD; PRT; 1152 AA.  
 ID CD45\_MOUSE  
 AC P05800;  
 DT 01-JAN-1988 (Rel. 06, Created)  
 DI 01-JAN-1988 (Rel. 06, Last sequence update)  
 DE 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE LEUCOCYTE COMMON ANTIGEN PRECURSOR (EC 3.1.3.48) (L-CA) (LYMPHOCYTE  
 COMMON ANTIGEN LY-5) (CD45) (T200).  
 GN PTERC OR LY-5.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 CC NCBI\_TaxID=10090;  
 FL [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=86313686; PubMed=2944116;  
 RA Saga Y., Tung J.-S., Shen F.-W., Boyse E.A.;  
 RT "Sequences of Ly-5 cDNA: Isoform-related diversity of Ly-5 mRNA.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 83:6940-6944(1986).  
 RN [2]  
 RP REVISIONS.  
 RA Saga Y., Tung J.-S., Shen F.-W., Boyse E.A.;  
 RL Proc. Natl. Acad. Sci. U.S.A. 84:1991-1991(1987).  
 RN [3]  
 RP SEQUENCE OF 10-124 FROM N.A.  
 RC TISSUE=T-cell.  
 RX MEDLINE=86042665; PubMed=3864163;  
 RA Shen F.-W., Saga Y., Litman G., Freeman G., Tung J.-S., Cantor H.,  
 RT Boyse E.A.;  
 RL "Cloning of Ly-5 cDNA";  
 RN Proc. Natl. Acad. Sci. U.S.A. 82:7360-7363(1985).  
 RP [4]  
 RP SEQUENCE OF 822-1152 FROM N.A.  
 RX MEDLINE=87092355; PubMed=2948186;  
 RA Raschke W.C.;  
 RT "Cloned murine T200 (Ly-5) cDNA reveals multiple transcripts within  
 B- and T-lymphocyte lineages";  
 RL Proc. Natl. Acad. Sci. U.S.A. 84:161-165(1987).  
 CC -1- FUNCTION: REQUIRED FOR T-CELLS ACTIVATION THROUGH THE ANTIGEN  
 CC RECEPTOR. THE FIRST PTASE DOMAIN HAS ENZYMACTIC ACTIVITY, WHILE  
 CC THE SECOND ONE SEEMS TO AFFECT THE SUBSTRATE SPECIFICITY OF THE  
 CC FIRST ONE.

CC -1- CATALYTIC ACTIVITY: PROTEIN TYROSINE PHOSPHATE + H(2)O =  
 CC PROTEIN TYROSINE + ORTHOPHOSPHATE.  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
 CC -1- ALTERNATIVE PRODUCTS: A NUMBER OF ISOFORMS ARE PRODUCED BY  
 CC ALTERNATIVE SPLICING.  
 CC -1- DEVELOPMENTAL STAGE: EXPRESSION IS RESTRICTED TO THE HEMATOPOIETIC  
 CC COMPARTMENT OF DEVELOPMENT.  
 CC -1- PTM: HEAVILY N- AND O-GLYCOSYLATED.  
 CC -1- SIMILARITY: CONTAINS 2 PROTEIN-TYROSINE PHOSPHATASE DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; M14342; AAA39458.1; -;  
 CC EMBL; M11934; AAA39461.1; -;  
 CC EMBL; M15174; AAA40161.1; -;  
 CC PIR; A29381; A29381.  
 CC HSP; P18052; 1YFO.  
 CC MGD; MGI:97810; Ptprc.  
 CC InterPro: IPR001777; FN\_III.  
 CC InterPro: IPR000387; Tyr\_phosphatase.  
 CC InterPro: IPR000242; Tyr\_prot\_phptase.  
 CC Pfam; PF00041; fn3; 3.  
 CC Pfam; PF00102; Y\_phosphatase; 2.  
 CC SMART; SM00194; PTPC; 2.  
 CC PRINTS; PR00700; PRTYPHPTASE.  
 CC PROSITE; PS00383; TYR\_PHOSPHATASE\_1; 2.  
 CC PROSITE; PS00056; TYR\_PHOSPHATASE\_2; 2.  
 CC PROSITE; PS00055; TYR\_PHOSPHATASE\_PTP; 2.  
 CC Glycoprotein; Transmembrane; Phosphorylation; B-cell; T-cell; Repeat;  
 CC Alternative splicing; Hydrolase; Signal.  
 CC SIGNAL 1 23  
 CC CHAIN 24 1152 LEUKOCYTE COMMON ANTIGEN.  
 CC DOMAIN 24 425 EXTRACELLULAR (POTENTIAL).  
 CC TRANSMEM 426 447 POTENTIAL.  
 CC DOMAIN 448 1152 CYTOPLASMIC (POTENTIAL).  
 CC DOMAIN 233 329 FIBRONECTIN TYPE-III 1.  
 CC DOMAIN 330 421 FIBRONECTIN TYPE-III 2.  
 CC DOMAIN 520 769 PROTEIN-TYROSINE PHOSPHATASE 1.  
 CC DOMAIN 811 1084 PROTEIN-TYROSINE PHOSPHATASE 2.  
 CC ACT\_SITE 701 701 BY SIMILARITY.  
 CC ACT\_SITE 1016 1016 BY SIMILARITY.  
 CC CARBOHYD 68 68 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC CARBOHYD 72 72 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC CARBOHYD 79 79 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC CARBOHYD 114 114 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC CARBOHYD 119 119 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC CARBOHYD 151 151 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC CARBOHYD 172 172 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC CARBOHYD 183 183 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC CARBOHYD 208 208 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC CARBOHYD 277 277 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC CARBOHYD 288 288 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC CARBOHYD 318 318 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC CARBOHYD 350 350 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC CARBOHYD 379 379 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC SEQUENCE 1152 AA; 130421 MW; B4D956B4E32EA812 CRC64;  
 Query Match 35.0%; Score 235.5; DB 1; Length 1152;  
 Best Local Similarity 37.6%; Pred. No. 7.8e-18;  
 Matches 47; Conservative 17; Mismatches 58; Indels 3; Gaps 2;  
 Qy 1 DFGGMNENCNVAMITREIEGVIVKCCSWPVSLEKPLEFKHFHVLLENFQITQYVI 60  
 Db 583 DFWRMIEQATVIVMVTGCEGNNRNCABEYWPMSMEGTRAFKDIVVTINDHRCRDPYII 642

Qy 61 RIFQIV-KKSTGKSHSVKHLQFIKWPHGTTPASVDFIKYVYR--KSHITGELLVHCT 117  
 Db 643 QKLVNAHKKEKATGREVTHIQFTSWPDHGVEDPHLLKLRRVNAFNSFSGPIVHCS 702  
 Qy 118 AGVGR 122  
 Db 703 AGVGR 707  
 RESULT 9  
 PTPM\_MOUSE  
 ID PTPM\_MOUSE STANDARD; PRT; 1452 AA.  
 AC P28828;  
 DT 01-DEC-1992 (Rel. 24, Created)  
 DT 01-DEC-1992 (Rel. 24, Last sequence update)  
 DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE PROTEIN-TYROSINE PHOSPHATASE MU PRECURSOR (EC 3.1.3.48) (R-PTP-MU).  
 GN PTPRM.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1].  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Lung;  
 RX MEDLINE=92008644; PubMed=1655529;  
 RA Gebbink M.F.B.G., van Etten I., Hateboer G., Suijkerbuijk R.,  
 RA Beijersbergen R., Geurts van Kessel A., Moolenaar W.H.;  
 PA "Cloning, expression and chromosomal localization of a new putative  
 PT receptor-like protein tyrosine phosphatase.";  
 NL FEBS Lett. 290:123-130(1991).  
 CC -1- FUNCTION: MAY PLAY A KEY ROLE IN SIGNAL TRANSDUCTION AND GROWTH  
 CC CONTROL.  
 CC -1- CATALYTIC ACTIVITY: PROTEIN TYROSINE PHOSPHATE + H(2)O =  
 CC PROTEIN TYROSINE + ORTHOPHOSPHATE.  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
 CC -1- TISSUE SPECIFICITY: MOST ABUNDANT IN LUNG, LESS IN BRAIN AND  
 CC HEART.  
 CC -1- SIMILARITY: CONTAINS 1 MAM DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.  
 CC -1- SIMILARITY: CONTAINS 4 FIBRONECTIN TYPE III-LIKE DOMAINS.  
 CC -1- SIMILARITY: CONTAINS 2 PROTEIN-TYROSINE PHOSPHATASE DOMAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; X58287; CAA41225.1; -;  
 CC PIR; S17670; S17670.  
 CC HSP; P28827; IRPM.  
 CC MGD; MGI:102694; Ptprn.  
 CC InterPro: IPR001777; FN\_III.  
 CC InterPro: IPR003599; Ig.  
 CC InterPro: IPR003006; Ig\_MHC.  
 CC InterPro: IPR000998; MAM.  
 CC InterPro: IPR000387; Tyr\_phosphatase.  
 CC InterPro: IPR000242; Tyr\_prot\_phptase.  
 CC Pfam; PF00041; fn3; 3.  
 CC Pfam; PF00047; Ig; 1.  
 CC Pfam; PF00629; MAM; 1.  
 CC Pfam; PF00102; Y\_phosphatase; 2.  
 CC PRINTS; PR00014; FNTYPEIII.  
 CC PRINTS; PR00020; MAMDOMAIN.  
 CC PRINTS; PR00700; PRTYPHPTASE.  
 CC SMART; SM00060; FN3; 2.  
 CC SMART; SM00409; IG; 1.  
 CC SMART; SM00137; MAM; 1.  
 CC SMART; SM00194; PTPC; 2.  
 CC PROSITE; PS00383; TYR\_PHOSPHATASE\_1; 2.

DR PROSITE: PS50056; TYR\_PHOSPHATASE\_2; 2.  
DR PROSITE: PS50055; TYR\_PHOSPHATASE\_PTP; 2.  
DR PROSITE: PS00740; MAM\_1; 1.  
DR PROSITE: PS50060; MAM\_2; 1.  
KW Hydrolase; Receptor; Glycoprotein; Signal; Transmembrane;  
KW Immunoglobulin domain; Repeat.  
FT SIGNAL 1 20  
FT CHAIN 21 1452  
FT DOMAIN 21 1452  
FT TRANSF 743 762  
FT DOMAIN 743 762  
FT DOMAIN 765 1452  
FT DOMAIN 22 184  
FT DOMAIN 199 267  
FT DOMAIN 287 374  
FT DOMAIN 383 466  
FT DOMAIN 486 571  
FT DOMAIN 589 671  
FT DOMAIN 923 1153  
FT DOMAIN 1213 1447  
FT ACT\_SITE 1095 1095  
FT ACT\_SITE 1389 1389  
FT DISULFID 206 260  
FT CARBOHYD 72 72  
FT CARBOHYD 92 92  
FT CARBOHYD 131 131  
FT CARBOHYD 249 249  
FT CARBOHYD 406 406  
FT CARBOHYD 414 414  
FT CARBOHYD 454 454  
FT CARBOHYD 534 534  
FT CARBOHYD 544 544  
FT CARBOHYD 598 598  
FT CARBOHYD 651 651  
FT CARBOHYD 681 681  
SQ SEQUENCE 1452 AA; 163594 MW; F99D0FC67922CF1E CRC64;  
Query Match 35.0%; Score 235.5; DB 1; Length 1452;  
Best Local Similarity 36.3%; Pred. No. le-17;  
Matches 45; Conservative 20; Mismatches 54; Indels 5; Gaps 2;  
QY 1 DFWGMWNNCNVIAITREIEGGVVKCSYWPVSLKEPFEKHFHVLLENFOITQYFVI 60  
DB 981 DFWRMVWHNTASIIWVNLVEGRVKCKYWP---DTEIYKDKIVLIDTELLAEYVI 1037  
QY 61 RIFQIVKSTGSKSHVKHLQFTKWDHGTDPASVDFFIKYVVRVRSKH--ITGPLLHVCTA 118  
DB 1038 RTFAVEKRGHIEIREIROFHFTGWDHGVYPYHATGLLGFVRQVSKSPNAGPLVVHCSA 1097  
QY 119 GVGR 122  
DB 1098 GAGR 1101  
RESULT 10  
CD45\_RAT  
ID CD45\_RAT STANDARD; PRT; 1255 AA.  
AC P04157;  
DT 01-NOV-1986 (Rel. 08, Created)  
DT 01-AUG-1988 (Rel. 03, Last sequence update)  
DT 20-AUG-2001 (Rel. 40, Last annotation update)  
DE LEUKOCYTE COMMON ANTIGEN PRECURSOR (EC 3.1.3.48) (L-CA) (CD45) (T200)  
DE (FRAGMENT).  
GN PTPRC.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
[1]  
RP SEQUENCE FROM N.A.  
RL Barclay A.N., Jackson D.I., Willis A.C., Williams A.F.;  
RL Submitted (MAY-1987) to the EMBL/GenBank/DBJ databases.  
RN [2]

SEQUENCE OF 190-1255 FROM N.A.  
MEDLINE-85201691; PubMed-3158393;  
RA Thomas M.L., Barclay A.N., Gagnon J., Williams A.F.;  
FT "Evidence from cDNA clones that the rat leukocyte-common antigen  
805000 Mr.";  
RT Cell 41:83-93(1985).  
RN [3];  
RP ALTERNATIVE SPLICING.  
RX MEDLINE-87275817; PubMed-2440674;  
RA Barclay A.N., Jackson D.I., Willis A.C., Williams A.F.;  
FT "Lymphocyte specific heterogeneity in the rat leukocyte common  
antigen (T200) is due to differences in polypeptide sequences near  
the NH2-terminus";  
RT EMBO J. 6:1259-1264(1987).  
CC -!- FUNCTION: REQUIRED FOR T-CELLS ACTIVATION THROUGH THE ANTIGEN  
RECEPTOR. THE FIRST PTPASE DOMAIN HAS ENZYMAIC ACTIVITY, WHILE  
THE SECOND ONE SEEMS TO AFFECT THE SUBSTRATE SPECIFICITY OF THE  
FIRST ONE.  
CC -!- CATALYTIC ACTIVITY: PROTEIN TYROSINE PHOSPHATASE + H(2)O =  
PROTEIN TYROSINE + ORTHOPHOSPHATE.  
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.  
CC -!- ALTERNATIVE PRODUCTS: THERE ARE DIFFERENT VARIANTS OF L-CA, WHICH  
ARISE BY ALTERNATIVE SPLICING.  
CC -!- TISSUE SPECIFICITY: VARIANTS 4 AND 3 ARE FOUND IN THE LYMPH NODE,  
VARIANTS 1 AND 2 ARE FOUND IN THYMOCYTE AND LYMPH NODE.  
CC -!- PTM: HEAVILY N- AND O-GLYCOSYLATED.  
CC -!- PTM: THE CYTOPLASMIC DOMAIN CONTAINS POTENTIAL PHOSPHORYLATION  
SITES.  
CC -!- SIMILARITY: CONTAINS 2 PROTEIN-TYROSINE PHOSPHATASE DOMAINS.  
CC -!- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.  
This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
DR EMBL; Y00065; CAA68272.1; -  
DR EMBL; Y00065; CAA68273.1; -  
DR EMBL; Y00065; CAA68274.1; -  
DR EMBL; Y00065; CAA68275.1; -  
DR EMBL; M25820; AAA41518.1; -  
DR EMBL; M25821; AAA41519.1; -  
DR EMBL; M25822; AAA41520.1; -  
DR EMBL; M25823; AAA41521.1; -  
DR PIR; A60241; TDRTLT.  
DR HSP; P18052; LYFO.  
DR InterPro; IPR001777; FN\_III.  
DR InterPro; IPR000387; TYR\_phosphatase.  
DR InterPro; IPR000242; Tyr\_prot\_phptase.  
DR Pfam; PF00041; fn3; 2.  
DR Pfam; PF00102; Y\_phosphatase; 2.  
DR SMART; SM00060; FN3; 2.  
DR SMART; SM00194; PTPC; 2.  
DR PROSITE; PS00383; TYR\_PHOSPHATASE\_1; 2.  
DR PROSITE; PS00056; TYR\_PHOSPHATASE\_2; 2.  
DR PROSITE; PS50055; TYR\_PHOSPHATASE\_PTP; 2.  
KW Glycoprotein; Transmembrane; Phosphorylation; B-cell; T-cell; Repeat;  
KW Alternative splicing; Hydrolase; Signal.  
FT NOBTER 1 1  
FT SIGNAL <1 5  
FT CHAIN 6 1255  
FT DOMAIN 6 528  
FT TRANSF 529 550  
FT DOMAIN 551 1255  
FT DOMAIN 341 432  
FT DOMAIN 433 524  
FT DOMAIN 623 872  
FT DOMAIN 914 1187  
FT ACT\_SITE 804 804  
L-CA, VARIANT 4.  
EXTRACELLULAR (POTENTIAL).  
CYTOPLASMIC (POTENTIAL).  
FIBRONECTIN TYPE-III 1.  
FIBRONECTIN TYPE-III 2.  
PROTEIN-TYROSINE PHOSPHATASE 1.  
PROTEIN-TYROSINE PHOSPHATASE 2.  
BY SIMILARITY.

RESULT	11
ID	PTNE_HUMAN
ID	PTNE_STANDARD; PRT; 1187 AA.
AC	Q15678;
DC	01-NOV-1997 (Rel. 35, Created)
DT	01-NOV-1997 (Rel. 35, Last sequence update)
DT	20-AUG-2001 (Rel. 40, Last annotation update)
DE	PROTEIN TYROSINE PHOSPHATASE, NON-RECEPTOR TYPE 14 (EC 3.1.3.48)
DE	(PROTEIN-TYROSINE PHOSPHATASE PEZ).
GN	PTPN14 OR PEZ.
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX	NCBI_TaxID=9606;
[1]	
SEQUENCE FROM N.A.	
RP	TISSUE-Breast carcinoma;
RX	MEDLINE=95251727; PubMed=7733990;
RA	Smith A.L., Mitchell P.J., Shipley J., Gusterson B.A., Rogers M.V.,
RA	Crompton M.R.;
RT	"Pez": a novel human cDNA encoding protein tyrosine phosphatase- and
RT	ezrin-like domains.";
RL	Biochem. Biophys. Res. Commun. 209:959-965(1995).
CC	-!- CATALYTIC ACTIVITY: PROTEIN TYROSINE PHOSPHATE + H(2)O =
CC	PROTEIN TYROSINE + ORTHOPHOSPHATE
CC	-!- TISSUE SPECIFICITY: EXPRESSED IN A VARIETY OF HUMAN TISSUES
CC	INCLUDING KIDNEY, SKELETAL MUSCLE, LUNG AND PLACENTA.
CC	-!- SIMILARITY: CONTAINS 1 BAND 4.1-LIKE DOMAIN.
CC	-!- SIMILARITY: BELONGS TO THE NON-RECEPTOR CLASS OF THE PROTEIN-
CC	TYROSINE PHOSPHATASE FAMILY.
CC	-----
CC	-----

RA Yamamoto T.;  
BT "Molecular cloning and characterization of Byp, a murine

RA	Yamamoto
12T	"Mol. Cryst. Liq. Cryst."

Molecular

"Molecular cloning and characterization of Byp, a murine



```

FT CARBOHYD 572 572 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 662 662 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 668 668 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 685 685 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 691 691 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 725 725 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 811 811 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 838 838 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 1238 AA; 136782 MW; 939479EDC8016835 CRC64;

Query: Match 34.5%; Score 232; DB 1; Length 1238;
Best Local Similarity 35.4%; Pred. No. 2e-17;
Matches 45; Conservative 23; Mismatches 51; Indels 8; Gaps 3;

QY 11 DFWGMWENNCNVIAMITREIEGGVIKCSYWPVSKLEPFEKHFHVLLNFOITQYFVI 60
DB 1023 DFWRMVWKNVYALVMTLKCEVGRKCEBYWP--SKQADYGDITVAMTSEVLPWTI 1080
QY 61 RFIQIVKKSSTGKSHSKVHLQFIKWPDHGTPASVDFFIKYVRYVKSHI-----TGPLLVH 115
DB 1081 RDEVVKMNQNSSEHPLRQPHFISWPDHGVDDTDLLINF-RYLVRDYMKQIPESPILVH 1139
QY 116 CTAGVGR 122
DB 1140 CSAGVGR 1146

RESULT: +13
FTPTD_HUMAN . STANDARD; PRT; 1912 AA.
AC P23468;
DE 01-NOV-1991 (Rel. 20, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE 20-AUG-2001 (Rel. 40, Last annotation update)
DE PRETEIN-TYROSINE PHOSPHATASE DELTA PRECURSOR (EC 3.1.3.48) (R-PTP-
DE DELTA).
DE GN DEGN
DE GN TPTPD.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC NCBI_TaxId:9606;
RN [1]
RX SEQUENCE FROM N.A., AND MUTAGENESIS OF ARG-1178.
RP MEDLINE=95204468; PubMed=7896816;
RA Pulido R., Krueger N.X., Serra-Pages C., Saito H., Streuli M.;
RT "Molecular characterization of the human transmembrane protein-
RT tyrosine phosphatase delta. Evidence for tissue-specific expression of
RT alternative human transmembrane protein-tyrosine phosphatase delta
RT isoforms."
RT J. Biol. Chem. 270:6722-6728(1995).
RN [2]
RP SEQUENCE OF 390-1912 FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=91006018; PubMed=2170109;
RA Krueger N.X., Streuli M., Saito H.;
RT "Structural diversity and evolution of human receptor-like protein
RT tyrosine phosphatases."
RL EM50 J. 9:3241-3252(1990).
CC 1- CATALYTIC ACTIVITY: PROTEIN TYROSINE PHOSPHATE + H(2)O =
CC PROTEIN TYROSINE + ORTHOPHOSPHATE.
CC 1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC 1- ALTERNATIVE PRODUCTS: DIFFERENT ISOFORMS ARE FOUND IN DIFFERENT
CC TISSUES DUE TO ALTERNATIVE SPLICING.
CC 1- PTM: A CLEAVAGE OCCURS THAT SEPARATES THE EXTRACELLULAR DOMAIN
CC FROM THE TRANSMEMBRANE SEGMENT.
CC 1- SIMILARITY: CONTAINS 3 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
CC 1- SIMILARITY: CONTAINS 8 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC 1- SIMILARITY: CONTAINS 2 PROTEIN-TYROSINE PHOSPHATASE DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation

```



```
FT DOMAIN 905 992 FIBRONECTIN TYPE-III 11.
FT DOMAIN 993 1082 FIBRONECTIN TYPE-III 12.
FT DOMAIN 1083 1170 FIBRONECTIN TYPE-III 13.
FT DOMAIN 1171 1268 FIBRONECTIN TYPE-III 14.
FT DOMAIN 1269 1352 FIBRONECTIN TYPE-III 15.
FT DOMAIN 1353 1442 FIBRONECTIN TYPE-III 16.
FT DOMAIN 1443 1530 FIBRONECTIN TYPE-III 17.
FT ACT_SITE 1531 1618 FIBRONECTIN TYPE-III 18.
FT CARBOHYD 28 28 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 53 53 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 75 75 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 172 172 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 198 198 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 267 267 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 321 321 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 414 414 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 421 421 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 479 479 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 544 544 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 574 574 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 598 598 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 652 652 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 721 721 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 829 829 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1040 1040 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1096 1096 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1163 1163 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1185 1185 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1212 1212 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1274 1274 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1367 1367 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1470 1470 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1474 1474 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1518 1518 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1519 AA; 224267 MW; 691E998A7A1515DD CRC64;

Query Match 34.3%; Score 231; DB 1; Length 1997;
Best Local Similarity 35.4%; Pred. No. 4.4e-17;
Matches 45; Conservative 27; Mismatches 49; Indels 6; Gaps 3;

QY 1 DFQGMWENNCNVIAMITREIEGGVVKCSYWPVSLKEPFEKHFVLELNFQITQYFVI 60
DQ 1785 DFQGMWENNCNVIAMITREIEGGVVKCSYWPVSLKEPFEKHFVLELNFQITQYFVI 1843

QY 61 RIFQIVKSTGKSHS-VKHLQFIKWPDHGTPTASVDFFIKYR-----YVRKSHITGPLLHV 115
DQ 1844 REFKICGERQLDAHLIRHFHTVWPDHGVPTTQSLIQFVTRVDRYINRSPGAGPTVVH 1903

QY 116 CTAGVGR 122
DQ 1904 CSAGVGR 1910

RESULT 15
CD45_HUMAN
ID CD45_HUMAN STANDARD: PRT: 1304 AA.
AC P08575;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE LEUKOCYTE COMMON ANTIGEN PRECURSOR (EC 3.1.1.3.48) (L-CA) (CD45 ANTIGEN) (T200).
DE (T200).
GN PTPRC OR CD45.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lymphocytes;
RX MEDLINE=88061067; PubMed=2824653;
RA Streuli M., Hall L.R., Saga Y., Schlossman S.F., Saito H.;
```

"Differential usage of three exons generates at least five different mRNAs encoding human leukocyte common antigens.";

J. Exp. Med. 166:1548-1566(1987).

[2]

RN FUNCTION.

RX MEDLINE=89017162; PubMed=2845400;

RA Charbonneau H., Tonks N.K., Walsh K.A., Fischer E.H.;

RT "The leukocyte common antigen (CD45): a putative receptor-linked protein tyrosine phosphatase.";

RT Proc. Natl. Acad. Sci. U.S.A. 85:7182-7186(1988).

RN [3]

RX MUTAGENESIS.

RA MEDLINE=90316093; PubMed=1895146;

RT Streuli M., Krueger N.X., Thal T., Tang M., Saito H.;

RT "Distinct functional roles of the two intracellular phosphatase like domains of the receptor-linked protein tyrosine phosphatases LCA and LAR.";

RL EMBO J. 9:2399-2407(1990).

CC 1-15 FUNCTION: REQUIRED FOR T-CELLS ACTIVATION THROUGH THE ANTIGEN RECEPTOR. THE FIRST PTPASE DOMAIN HAS ENZYMIC ACTIVITY, WHILE THE SECOND ONE SEEMS TO AFFECT THE SUBSTRATE SPECIFICITY OF THE FIRST ONE.

CC 1-15 CATALYTIC ACTIVITY: PROTEIN TYROSINE PHOSPHATASE + H(2)O = SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.

CC 1-15 ALTERNATIVE PRODUCTS: AT LEAST 8 ISOFORMS ARE PRODUCED BY ALTERNATIVE SPLICING.

CC 1-15 PTM: HEAVILY N- AND O-GLYCOSYLATED.

CC 1-15 SIMILARITY: CONTAINS 2 PROTEIN-TYROSINE PHOSPHATASE DOMAINS.

CC 1-15 SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.

CC 1-15 DATABASE: NAME=PROW; NOTE=CD guide CD45 entry;

CC 1-15 WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cd45.htm".

CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DQ EMAL: Y00638; CAA68669.1; .

DQ HSP: P18052; IYEO.

DQ GlycoSuitDB; P08575; .

DQ MIM: 151460; .

DQ InterPro: IPR001777; FN\_III.

DQ InterPro: IPR000387; TVR\_phosphatase.

DQ InterPro: IPR000242; Tyr\_prot\_phptase.

DQ Pfam: PF00041; fn3; 2.

DQ Pfam: PF00102; Y\_phosphatase; 2.

DQ PRINTS: PR00700; PRTYPHPTASE.

DQ SMART: SM00060; FN3; 2.

DQ SMART: SM00194; PTPc; 2.

DQ PROSITE: PS00383; TYR\_PHOSPHATASE\_1; 2.

DQ PROSITE: PS00386; TYR\_PHOSPHATASE\_2; 2.

DQ PROSITE: PS00055; TYR\_PHOSPHATASE\_PTP; 2.

DQ Glycoprotein; Transmembrane; Phosphorylation; B-cell; T-cell; Repeat; Alternative splicing; Hydrolase; Signal.

DQ SIGNAL 1 23

FT CHAIN 24 1304 LEUKOCYTE COMMON ANTIGEN.

FT DOMAIN 24 575 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 576 597 POTENTIAL.

FT DOMAIN 598 1304 CYTOPLASMIC (POTENTIAL).

FT DOMAIN 387 479 FIBRONECTIN TYPE-III 1.

FT DOMAIN 480 571 FIBRONECTIN TYPE-III 2.

FT DOMAIN 670 919 PROTEIN-TYROSINE PHOSPHATASE 1.

FT DOMAIN 961 1235 PROTEIN-TYROSINE PHOSPHATASE 2.

FT ACT\_SITE 851 851

FT ACET\_SITE 1167 1167

FT CARBOHYD 78 78 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 90 90 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 95 95 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 184 184 N-LINKED (GLCNAC. . .) (POTENTIAL).

```
FT CARBOHYD 190 190 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 197 197 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 232 232 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 260 260 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 270 270 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 276 276 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 335 335 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 378 378 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 419 419 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 468 468 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 488 488 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 529 529 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT MUTAGEN 851 851 C->S: LOSS OF ACTIVITY.
SQ SEQUENCE 1304 AA: 147253 MW: 1F357EC5632618B2 CRC64;

Query Match 34.2%; Score 230.5; DB 1; Length 1304;
Best Local Similarity 38.4%; Pred. No. 3.1e-17;
Matches 48; Conservative 15; Mismatches 59; Indels 3; Gaps 2;

QY 1 DFWGMWNNCNIAMITREIEGGVINKCCSYWPVSLKEPLEFKHFHVLLLENFOITQYFVI 60
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 733 DFWRMIEQKATVIVMYTRCEEGNRNKAERYWPSMEEGTAFGVDVVVKINQHRKCPDYII 792

QY 61 RIFQIV-KKSTGKSHSVKHLQFIKWPDPHGTGPASVDPEFIKYRVYVR--KSHITGPLLHCT 117
   || || || || || || || || || || || || || || || || || || || || || ||
Db 793 OKLNIYVKKKATGREVTHIQFTSWDPDHGVDPDPLLKLRRRVNAFSPFFSGPIVVHCS 852

QY 118 AGVGR 122
   |||||
Db 853 AGVGR 857
```

Search completed: March 30, 2002, 08:26:05  
Job time: 657 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 30, 2002, 07:17:38 ; Search time 69.17 Seconds  
(without alignments)  
134.354 Million cell updates/sec

Title: US-09-095-478A-8  
Perfect score: 673  
Sequence: 1 DFWGMWENNVCNVIAMITRE.....VRKSHITGPLLHVCTAGVGR 122  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues  
Total number of hits satisfying chosen parameters: 219241  
Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR.68.\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result, being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	529	78.6	398	2	T08716	protein-tyrosine-p
2	274.5	40.8	2450	2	S71625	protein-tyrosine-p
3	268.5	39.9	2294	2	I67630	protein tyrosine p
4	268.5	39.9	2466	2	I67629	protein tyrosine p
5	266.5	39.6	377	1	A48711	protein-tyrosine-p
6	264.5	39.3	2490	1	A54971	protein-tyrosine-p
7	261	38.8	1156	2	T23308	hypothetical prote
8	250	37.1	913	1	A41109	protein-tyrosine-p
9	244	36.3	256	2	A40169	protein-tyrosine-p
10	240.5	35.7	597	2	B53978	protein-tyrosine-p
11	239	35.5	583	2	S17671	protein-tyrosine-p
12	239	35.5	1337	1	I38670	protein-tyrosine-p
13	238	35.4	2314	1	A46151	protein-tyrosine-p
14	237	35.2	1499	2	I50212	protein-tyrosine-p
15	236.5	35.1	1452	1	S17669	protein-tyrosine-p
16	235.5	35.0	1291	1	A28334	protein-tyrosine-p
17	235.5	35.0	1452	1	S17670	protein-tyrosine-p
18	233.5	34.7	694	2	A53978	protein-tyrosine-p
19	233.5	34.7	1273	1	TDR1T	leukocyte common a
20	232.5	34.5	1436	2	JC5230	protein-tyrosine-p
21	232	34.5	1187	1	JC4155	protein-tyrosine-p
22	232	34.5	1238	2	S68700	HTPP beta-like tyr
23	232	34.5	1912	2	A56178	protein-tyrosine-p
24	231	34.3	1501	2	I58148	protein-tyrosine-p
25	231	34.3	1907	2	S50893	protein-tyrosine-p
26	231	34.3	1997	1	S12050	protein-tyrosine-p
27	230.5	34.2	1304	1	A46546	leukocyte common a
28	230	34.2	582	2	A57068	protein-tyrosine p
29	230	34.2	1290	2	A56493	leucocyte common a

30	230	34.2	1897	1	TDHULK	leukocyte antigen-
31	230	34.2	1898	2	S46216	leukocyte antigen-
32	229.5	34.1	382	1	A38191	protein-tyrosine-p
33	229	34.0	2029	1	TDFFLK	protein-tyrosine-p
34	228.5	34.0	624	2	T19630	hypothetical prote
35	228.5	34.0	1026	2	T19631	hypothetical prote
36	228.5	34.0	1200	2	T43148	probable protein-t
37	228	33.9	2051	2	T30938	receptor tyrosine
38	227	33.7	1262	1	B48758	protein-tyrosine-p
39	227	33.7	1496	1	A48758	protein-tyrosine-p
40	227	33.7	1863	2	S46217	protein-tyrosine-p
41	226	33.6	382	1	S48748	protein-tyrosine-p
42	226	33.6	775	2	S55345	protein-tyrosine-p
43	226	33.6	1437	2	T31093	probable protein-t
44	225.5	33.5	363	1	S14294	protein-tyrosine-p
45	224.5	33.4	387	1	A60345	protein-tyrosine-p

ALIGNMENTS

RESULT 1

T08716

protein-tyrosine-phosphatase homolog DKF2p566K0524.1 - human (fragment)

C:Species: Homo sapiens (man)

C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 13-Aug-1999

C:Accession: T08716

R:Ansorge, W.; Winkner, U.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.

submitted to the Protein Sequence Database, May 1999

A:Reference number: Z16472

A:Accession: T08716

A:Molecule type: mRNA

A:Residues: 1-398 <ANS>

A:Cross-references: EMBL:AL050040

A:Experimental source: fetal kidney; clone DKF2p566K0524

C:Genetics:

A:Note: DKF2p566K0524.1

C:Superfamily: protein-tyrosine-phosphatase homology

F:161-379/Domain: protein-tyrosine-phosphatase homology <PTP>

Query Match.		78.6%;	Score 529;	DB 2;	Length 398;
Best Local Similarity		74.6%;	Pred. No. 2.3e-48;		
Matches		91;	Conservative	16;	Mismatches 15;
				Indels	0;
				Gaps	0;
Qy	{ 1	DFWGMWENNVCNVIAMITREIEGGVKKCSYWPVSLKEPFEKHFHVLLENFQITQYFVI	60		

RESULT 2

S71625

protein-tyrosine-phosphatase (EC 3.1.3.48) RIP - mouse

N:Alternate names: epidermal growth factor-binding protein; serine proteinase

C:Species: Mus musculus (house mouse)

C:Date: 27-Nov-1997 #sequence\_revision 12-Dec-1997 #text\_change 20-Jun-2000

C:Accession: S71625; S67987

R:Chida; D.; Kume, T.; Mukoyama, Y.; Tabata, S.; Nomura, N.; Thomas, M.L.; Watanabe,

FEBS Lett. 358, 233-239, 1995

A:Title: Characterization of a protein tyrosine phosphatase (RIP) expressed at a very

A:Reference number: S71625; MUID:95145716

A:Accession: S71625

A:Molecule type: mRNA

A:Residues: 1-2450 <CHI>

A:Cross-references: EMBL:D83966; NID:g1232103; PIDN:BAA12158.1; PID:g1232104

A:Experimental source: strain DBA/2; cell line MEL 745A

R:Wolf, B.S.; Brown, M.D., 1995

F:1089-1165/Domain: GLGF domain homology <GLG1>

A:Title: Epidermal growth factor-binding protein activates soluble and receptor-bound signal transduction

A:Reference number: S67987; MUID:96105375

A:Accession: S67987

A:Molecule type: protein

A:Residues: 1098-1102 <MOL>

A:Experimental source: submaxillary glands

C:Genetics:

A:Gene: Ptpn13

A:Map position: 5

C:Superfamily: protein-tyrosine-phosphatase, nonreceptor type 13; GLGF domain homology; F:566-860/Domain: protein 4.1 membrane-binding domain homology <B41>

F:1089-1165/Domain: GLGF domain homology <GLG1>

F:1361-1437/Domain: GLGF domain homology <GLG2>

F:1495-1574/Domain: GLGF domain homology <GLG3>

F:1769-1840/Domain: GLGF domain homology <GLG4>

F:1863-1937/Domain: GLGF domain homology <GLG5>

F:2203-2422/Domain: protein-tyrosine-phosphatase homology <PTP1>

F:2374/Active site: Cys (phosphotyrosine intermediate) #status predicted

F:2380/Binding site: substrate phosphate (Arg) #status predicted

Query Match 40.8%; Score 274.5; DB 2; Length 2450;

Best Local Similarity 41.3%; Pred. No. 1.4e-20;

Matches 52; Conservative 23; Mismatches 44; Indels 7; Gaps 3;

QY 1 DFWGMWNNCNCVIAITREIEGVKCCSWPVSL-KEPLEKHFHVLLENFOITQYFV 59

DB 2258 DFWQMIWEQKSTVIAMTQVEGEKIKQRYWPNILGKTTMWSNRLALVRMQQLKGFV 2317

QY 60 IRIFOIVKKGSTGSHSVKHLQFTKPDHGTSPASVD---FFIKYRVYRKSHITGPLLHVC 116

DB 2318 RVYMALEDIQTEVGRHSHLNFTAMPDHTPSQPDLLTFTSYMRHRS---GPVITHC 2374

QY 117 TAGVGR 122

DB 2375 SAGIGR 2380

RESULT 3

I67630

protein tyrosine phosphatase (PTP-BAS, type 3) - human

C:Species: Homo sapiens (man)

C:Date: 29-May-1998 #sequence\_revision 29-May-1998 #text\_change 16-Jul-1999

C:Accession: I67630

R:Maekawa, K.; Imagawa, N.; Nagamatsu, M.; Harada, S.

F:1089-1165/Domain: GLGF domain homology <GLG1>

F:1361-1437/Domain: GLGF domain homology <GLG2>

F:1495-1574/Domain: GLGF domain homology <GLG3>

F:1769-1840/Domain: GLGF domain homology <GLG4>

F:1863-1937/Domain: GLGF domain homology <GLG5>

F:2203-2422/Domain: protein-tyrosine-phosphatase homology <PTP1>

F:2374/Active site: Cys (phosphotyrosine intermediate) #status predicted

F:2380/Binding site: substrate phosphate (Arg) #status predicted

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-2294 <RES>

C:Cross-references: GB:D21211; NID:g452193; PIDN:BAA04752.1; PID:g452194

C:Superfamily: protein-tyrosine-phosphatase, nonreceptor type 13; GLGF domain homology; F:574-868/Domain: protein 4.1 membrane-binding domain homology <B41>

F:1182-1258/Domain: GLGF domain homology <GLG2>

F:2046-2265/Domain: protein-tyrosine-phosphatase homology <PTP>

Query Match 39.9%; Score 268.5; DB 2; Length 2294;

Best Local Similarity 39.8%; Pred. No. 5.6e-20;

Matches 49; Conservative 21; Mismatches 52; Indels 1; Gaps 1;

QY 1 DFWGMWNNCNCVIAITREIEGVKCCSWPVSL-KEPLEKHFHVLLENFOITQYFV 59

DB 2101 DFWQMIWEQKSTVIAMTQVEGEKIKQRYWPNILGKTTMWSNRLALVRMQQLKGFV 2160

QY 60 IRIFOIVKKGSTGSHSVKHLQFTKPDHGTSPASVDFFIKYRVYRKSHITGPLLHVC 119

DB 2375 SAGIGR 2380

RESULT 3

I67630

protein tyrosine phosphatase (PTP-BAS, type 3) - human

C:Species: Homo sapiens (man)

C:Date: 29-May-1998 #sequence\_revision 29-May-1998 #text\_change 16-Jul-1999

C:Accession: I67630

R:Maekawa, K.; Imagawa, N.; Nagamatsu, M.; Harada, S.

F:1089-1165/Domain: GLGF domain homology <GLG1>

F:1361-1437/Domain: GLGF domain homology <GLG2>

F:1495-1574/Domain: GLGF domain homology <GLG3>

F:1769-1840/Domain: GLGF domain homology <GLG4>

F:1863-1937/Domain: GLGF domain homology <GLG5>

F:2203-2422/Domain: protein-tyrosine-phosphatase homology <PTP1>

F:2374/Active site: Cys (phosphotyrosine intermediate) #status predicted

F:2380/Binding site: substrate phosphate (Arg) #status predicted

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-2294 <RES>

C:Cross-references: GB:D21211; NID:g452193; PIDN:BAA04752.1; PID:g452194

C:Superfamily: protein-tyrosine-phosphatase, nonreceptor type 13; GLGF domain homology; F:574-868/Domain: protein 4.1 membrane-binding domain homology <B41>

F:1182-1258/Domain: GLGF domain homology <GLG2>

F:2046-2265/Domain: protein-tyrosine-phosphatase homology <PTP>

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-2294 <RES>

C:Cross-references: GB:D21211; NID:g452193; PIDN:BAA04752.1; PID:g452194

C:Superfamily: protein-tyrosine-phosphatase, nonreceptor type 13; GLGF domain homology; F:574-868/Domain: protein 4.1 membrane-binding domain homology <B41>

F:1182-1258/Domain: GLGF domain homology <GLG2>

F:2046-2265/Domain: protein-tyrosine-phosphatase homology <PTP>

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-2294 <RES>

C:Cross-references: GB:D21211; NID:g452193; PIDN:BAA04752.1; PID:g452194

C:Superfamily: protein-tyrosine-phosphatase, nonreceptor type 13; GLGF domain homology; F:574-868/Domain: protein 4.1 membrane-binding domain homology <B41>

F:1182-1258/Domain: GLGF domain homology <GLG2>

F:2046-2265/Domain: protein-tyrosine-phosphatase homology <PTP>

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-2294 <RES>

C:Cross-references: GB:D21211; NID:g452193; PIDN:BAA04752.1; PID:g452194

C:Superfamily: protein-tyrosine-phosphatase, nonreceptor type 13; GLGF domain homology; F:574-868/Domain: protein 4.1 membrane-binding domain homology <B41>

F:1182-1258/Domain: GLGF domain homology <GLG2>

F:2046-2265/Domain: protein-tyrosine-phosphatase homology <PTP>

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-2294 <RES>

C:Cross-references: GB:D21211; NID:g452193; PIDN:BAA04752.1; PID:g452194

C:Superfamily: protein-tyrosine-phosphatase, nonreceptor type 13; GLGF domain homology; F:574-868/Domain: protein 4.1 membrane-binding domain homology <B41>

F:1182-1258/Domain: GLGF domain homology <GLG2>

F:2046-2265/Domain: protein-tyrosine-phosphatase homology <PTP>

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-2294 <RES>

C:Cross-references: GB:D21211; NID:g452193; PIDN:BAA04752.1; PID:g452194

C:Superfamily: protein-tyrosine-phosphatase, nonreceptor type 13; GLGF domain homology; F:574-868/Domain: protein 4.1 membrane-binding domain homology <B41>

F:1182-1258/Domain: GLGF domain homology <GLG2>

F:2046-2265/Domain: protein-tyrosine-phosphatase homology <PTP>

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-2294 <RES>

C:Cross-references: GB:D21211; NID:g452193; PIDN:BAA04752.1; PID:g452194

C:Superfamily: protein-tyrosine-phosphatase, nonreceptor type 13; GLGF domain homology; F:574-868/Domain: protein 4.1 membrane-binding domain homology <B41>

F:1182-1258/Domain: GLGF domain homology <GLG2>

F:2046-2265/Domain: protein-tyrosine-phosphatase homology <PTP>

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-2294 <RES>

C:Cross-references: GB:D21211; NID:g452193; PIDN:BAA04752.1; PID:g452194

C:Superfamily: protein-tyrosine-phosphatase, nonreceptor type 13; GLGF domain homology; F:574-868/Domain: protein 4.1 membrane-binding domain homology <B41>

F:1182-1258/Domain: GLGF domain homology <GLG2>

F:2046-2265/Domain: protein-tyrosine-phosphatase homology <PTP>

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-2294 <RES>

C:Cross-references: GB:D21211; NID:g452193; PIDN:BAA04752.1; PID:g452194

C:Superfamily: protein-tyrosine-phosphatase, nonreceptor type 13; GLGF domain homology; F:574-868/Domain: protein 4.1 membrane-binding domain homology <B41>

F:1182-1258/Domain: GLGF domain homology <GLG2>

F:2046-2265/Domain: protein-tyrosine-phosphatase homology <PTP>

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-2294 <RES>

C:Cross-references: GB:D21211; NID:g452193; PIDN:BAA04752.1; PID:g452194

C:Superfamily: protein-tyrosine-phosphatase, nonreceptor type 13; GLGF domain homology; F:574-868/Domain: protein 4.1 membrane-binding domain homology <B41>

F:1182-1258/Domain: GLGF domain homology <GLG2>

F:2046-2265/Domain: protein-tyrosine-phosphatase homology <PTP>

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-2294 <RES>

C:Cross-references: GB:D21211; NID:g452193; PIDN:BAA04752.1; PID:g452194

C:Superfamily: protein-tyrosine-phosphatase, nonreceptor type 13; GLGF domain homology; F:574-868/Domain: protein 4.1 membrane-binding domain homology <B41>

F:1182-1258/Domain: GLGF domain homology <GLG2>

F:2046-2265/Domain: protein-tyrosine-phosphatase homology <PTP>

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-2294 <RES>

C:Cross-references: GB:D21211; NID:g452193; PIDN:BAA04752.1; PID:g452194

C:Superfamily: protein-tyrosine-phosphatase, nonreceptor type 13; GLGF domain homology; F:574-868/Domain: protein 4.1 membrane-binding domain homology <B41>

F:1182-1258/Domain: GLGF domain homology <GLG2>

F:2046-2265/Domain: protein-tyrosine-phosphatase homology <PTP>

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-2294 <RES>

C:Cross-references: GB:D21211; NID:g452193; PIDN:BAA04752.1; PID:g452194

C:Superfamily: protein-tyrosine-phosphatase, nonreceptor type 13; GLGF domain homology; F:574-868/Domain: protein 4.1 membrane-binding domain homology <B41>

F:1182-1258/Domain: GLGF domain homology <GLG2>

F:2046-2265/Domain: protein-tyrosine-phosphatase homology <PTP>

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-2294 <RES>

C:Cross-references: GB:D21211; NID:g452193; PIDN:BAA04752.1; PID:g452194

C:Superfamily: protein-tyrosine-phosphatase, nonreceptor type 13; GLGF domain homology; F:574-868/Domain: protein 4.1 membrane-binding domain homology <B41>

F:1182-1258/Domain: GLGF domain homology <GLG2>

F:2046-2265/Domain: protein-tyrosine-phosphatase homology <PTP>

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-2294 <RES>

C:Cross-references: GB:D21211; NID:g452193; PIDN:BAA04752.1; PID:g452194

C:Superfamily: protein-tyrosine-phosphatase, nonreceptor type 13; GLGF domain homology; F:574-868/Domain: protein 4.1 membrane-binding domain homology <B41>

F:1182-1258/Domain: GLGF domain homology <GLG2>

F:2046-2265/Domain: protein-tyrosine-phosphatase homology <PTP>

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-2294 <RES>

C:Cross-references: GB:D21211; NID:g452193; PIDN:BAA04752.1; PID:g452194

C:Superfamily: protein-tyrosine-phosphatase, nonreceptor type 13; GLGF domain homology; F:574-868/Domain: protein 4.1 membrane-binding domain homology <B41>

F:1182-1258/Domain: GLGF domain homology <GLG2>

F:2046-2265/Domain: protein-tyrosine-phosphatase homology <PTP>

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-2294 <RES>

C:Cross-references: GB:D21211; NID:g452193; PIDN:BAA04752.1; PID:g452194

C:Superfamily: protein-tyrosine-phosphatase, nonreceptor type 13; GLGF domain homology; F:574-868/Domain: protein 4.1 membrane-binding domain homology <B41>

F:1182-1258/Domain: GLGF domain homology <GLG2>

F:2046-2265/Domain: protein-tyrosine-phosphatase homology <PTP>

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-2294 <RES>

C:Cross-references: GB:D21211; NID:g452193; PIDN:BAA04752.1; PID:g452194

C:Superfamily: protein-tyrosine-phosphatase, nonreceptor type 13; GLGF domain homology; F:574-868/Domain: protein 4.1 membrane-binding domain homology <B41>

F:1182-1258/Domain: GLGF domain homology <GLG2>

F:2046-2265/Domain: protein-tyrosine-phosphatase homology <PTP>

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-2294 <RES>

C:Cross-references: GB:D21211; NID:g452193; PIDN:BAA04752.1; PID:g452194

C:Superfamily: protein-tyrosine-phosphatase, nonreceptor type 13; GLGF domain homology; F:574-868/Domain: protein 4.1 membrane-binding domain homology <B41>

F:1182-1258/Domain: GLGF domain homology <GLG2>

F:2046-2265

F;1099-1175/Domain: GLGF domain homology <GLG1>  
F;1373-1454/Domain: GLGF domain homology <GLG2>  
F;1511-1590/Domain: GLGF domain homology <GLG3>

F:1099-1175/Domain: GLGF domain homology <GLG1>  
F:1373-1454/Domain: GLGF domain homology <GLG2>  
F:1511-1590/Domain: GLGF domain homology <GLG3>

```
C;Date: 27
C;Accessio
R;Yang, Q.
Proc. Natl
A;Title: In
A;Referenc
```

Proc. Natl. Acad. Sci. U.S.A. 88, 5949-5953, 1991  
A; title: Isolation of a cDNA clone encoding a human p  
A; reference number: A41109; MUID: 51296738

```
A:Accession: A41109
A:Molecule type: mRNA
A:Residues: 1-913 <YAN>
A:Cross-references: GB:M64572; NID:g179912; PIDN:AAA35647.1; PID:g179913
R:Ikuta, S.; Itoh, F.; Hindoda, Y.; Toyota, M.; Makiguchi, Y.; Imai, K.; Yachi, A.
J. Gastroenterol. 29, 727-732, 1994
A:title: Expression of cytoskeletal-associated protein tyrosine phosphatase PTPH1 mRNA in
A:Reference number: I55698; MUID:95179278
A:Accession: I55698
A:Status: translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 899-913 <RES>
A:Cross-references: GB:S76309; NID:g913165; PIDN:AAB33583.1; PID:g913166
C:Genetics:
A:Gene: GDB:PTPN3
A:Cross-references: GDB:I31386; OMIM:176877
A:Map position: 9q31-9q31
A:Superfamily: protein-tyrosine-phosphatase, nonreceptor type 3; GLGF domain homology; F
C:Keywords: phosphoprotein; protein-tyrosine-phosphatase; tyrosine-specific phosphatase
F:31-308/Domain: protein 4.1 membrane-binding domain homology <B41>
F:516-590/Domain: GLGF domain homology <GLG>
F:670-890/Domain: protein-tyrosine-phosphatase homology <PTP>
F:842/Active site: Cys (phosphocysteine intermediate) #status predicted
F:848/Binding site: substrate phosphate (Arg) #status predicted

Query Match 37.1%; Score 250; DB 1; Length 913;
Best Local Similarity 37.9%; Pred. No. 1.9e-18;
Matches 47; Conservative 23; Mismatches 48; Indels 6; Gaps 3;

QY 2 FWMGMWNNCNCVIAITREIEGVKCCSYWPVSLKEPFEKH--FHVLENFQITQYFV 59
Db 728 FQVQVMDKUSLIVMLTTLTERGTRKCHQYWP---DPDVMNHHGGFHQCOSEDTAYV 784

QY 60 IRIFOIVKSTGKSHSVKHLQFIKWPDHGTPASVDFFIKYVRYVRKSHI-TGPLLVHCTA 118
Db 785 SREMLVTNTQGEHTVTHLQYVAMPDHGIPDDSSDFLEFNVYRSLRVDSPEVLVHCSA 844

QY 119 GVGR 122
Db 845 GIGR 848

RESULT 9
A01069
protein-tyrosine-phosphatase (EC 3.1.3.48), receptor type zeta - rat (fragment)
N:Alternate names: protein-tyrosine-phosphatase beta; protein-tyrosine-phosphatase PTPB
C:Species: Rattus norvegicus (Norway rat)
C:Date: 05-Jun-1992 #sequence_revision 05-Jun-1992 #text_change 24-Apr-1998
C:Accession: A40169
R:Guan, K.; Dixon, J.E.
Science 249, 553-556, 1990
A:title: Protein tyrosine phosphatase activity of an essential virulence determinant in
A:Reference number: A40169; MUID:90341778
A:Accession: A40169
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-256 <GUA>
C:Function:
A:Description: may be involved in the regulation of specific developmental processes in
C:Superfamily: protein-tyrosine-phosphatase, receptor type zeta; carbonic anhydrase hom
C:Keywords: brain; glycoprotein; phosphoprotein; phosphoric monoester hydrolase; recep
F:6-237/Domain: protein-tyrosine-phosphatase homology <PP1>
F:189/Active site: Cys (phosphocysteine intermediate) #status predicted
F:195/Binding site: substrate phosphate (Arg) #status predicted

Query Match 36.3%; Score 244; DB 2; Length 256;
Best Local Similarity 38.3%; Pred. No. 2.1e-18;
Matches 52; Conservative 22; Mismatches 43; Indels 18; Gaps 5;

QY 1 DFWMGMWNNCNCVIAITREIEGVKCCSYWPVSLKEPFEKHFFHVLENFQITQYFVI 60
Db 386 DFWMGMWNNCNCVIAITREIEGVKCCSYWPVSLKEPFEKHFFHVLENFQITQYFVI 60

Query Match 35.7%; Score 240.5; DB 2; Length 597;
Best Local Similarity 35.0%; Pred. No. 1.2e-17;
Matches 48; Conservative 21; Mismatches 53; Indels 15; Gaps 2;

QY 1 DFWMGMWNNCNCVIAITREIEGVKCCSYWPVSLKEPFEKHFFHVLENFQITQYFVI 60
Db 386 DFWMGMWNNCNCVIAITREIEGVKCCSYWPVSLKEPFEKHFFHVLENFQITQYFVI 60

QY 61 RIFOIVKSTGKSHSVKHLQFIKWPDHGTPASVDFFIKYVRYVRK-----S 106
Db 446 THEEVYNKQDESRAHYQYMSWPDFGVKPSASAMLDERSQVHQVAVQNLGMEWTG 505

QY 107 HITG-PLLVHCTAGVGR 122
Db 506 HPAGPPIVVHCSAGIGR 522

RESULT 11
S17671
protein-tyrosine-phosphatase (EC 3.1.3.48), receptor type beta - mouse (fragment)
C:Species: Mus musculus (house mouse)
C:Date: 22-Nov-1993 #sequence_revision 15-Mar-1996 #text_change 23-Jul-1999
C:Accession: S17671; S40287
R:Gebink, M.F.B.G.; van Erten, I.; Hateboer, G.; Suijkerbuijk, R.; Beijersbergen, R.
FEBS Lett. 290, 123-130, 1991
A:title: Cloning, expression and chromosomal localization of a new putative receptor-
A:Reference number: S17669; MUID:92008644
A:Status: S17671
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: mRNA
A:Residues: 1-583 <GEB>
A:Cross-references: EMBL:X58289
R:Hendriks, W.; Brugman, C.; Zeeuwen, P.; Schepens, J.; Wieringa, B.
submitted to the EMBL Data Library, June 1993
A:Description: Assessment of the expression levels of murine protein-tyrosine phosph
A:Reference number: S40280
A:Accession: S40287
```



A:Molecule type: mRNA  
A:Residues: 377-483,'m',485-486 <HEN>  
A:Cross-references: EMBL:223056; NID:9438149; PIDN:CAA80591.1; PID:g438150  
C:Superfamily: protein-tyrosine-phosphatase, receptor type beta; fibronectin type III re  
C:Keywords: phosphoprotein; phosphoric monoester hydrolase; receptor; Transmembrane prot  
F:311-536/Domain: protein-tyrosine-phosphatase homology <PTP>  
F:488/Active site: Cys (phosphocysteine intermediate) #status predicted  
F:494/Binding site: substrate phosphate (Arg) #status predicted

Query Match 35.5%; Score 239; DB 2; Length 583;  
Best Local Similarity 36.2%; Pred. No. 1.7e-17;  
Matches 46; Conservative 26; Mismatches 49; Indels 6; Gaps 3;

QY 1 DFWGMMWNNCNIAMITREIEGGVKKCSYWPVSLKEPLEKHFHVLLENFOITQYFVI 60  
DB 369 DFWKMAEQNVHINIVMTQCEKGRVKCDHYPAD-QDPLYGDLTLQWVSESLPEWTI 427  
QY 61 RIFQIVKKSSTGKSHS-VKHLQFIKWPDHCTPASVDFFIKYV---YVRKSHITGPLLHV 115  
DB 428 REFKICSEQLDAHLIRHFHYTVMPDHGVPETTSLOFVTRVDYINRSPGAGPSVVH 487  
QY 116 CTAGVGR 122  
DB 488 CSAGVGR 494

RESULT 12  
138670  
protein-tyrosine-phosphatase (EC 3.1.3.48), receptor type J precursor - human  
N:Alternate names: density enhanced phosphatase-1 (DEP-1); protein-tyrosine phosphatase  
C:Species: Homo sapiens (man)  
C:Date: 01-Mar-1996 #sequence\_revision 08-Mar-1996 #text\_change 22-Jun-1999  
C:Accession: 138670; 152599  
R:Ostman, A.; Yang, Q.; Tonks, N.K.  
Proc. Natl. Acad. Sci. U.S.A. 91, 9680-9684, 1994  
A:Title: Expression of DEP-1, a receptor-like protein-tyrosine-phosphatase, is enhanced  
A:Reference number: 138670; MUID:95024024  
A:Accession: 138670  
A:Molecule type: mRNA  
A:Residues: 1-1337 <RES>  
A:Cross-references: EMBL:U10886; NID:g558754; PID:g558755  
A:Experimental source: HeLa cells  
R:Honda, H.; Inazawa, J.; Nishida, J.; Yazaki, Y.; Hirai, H.  
Blood 84, 4186-4194, 1994  
A:Title: Molecular cloning, characterization, and chromosomal localization of a novel pr  
A:Reference number: 152599; MUID:95086212  
A:Accession: 152599  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: mRNA  
A:Residues: 1-216,'LTGVRRAA',225-260,'G',262-285,'GTGGGLDASNTERRSRA',302,'S',304,'TAPVHDE  
A:Cross-references: GB:D37781; NID:g633072; PIDN:BAA07035.1; PID:g633073  
C:Comment: Enhanced expression of this protein with increasing cell density suggests a r  
C:Genetics:  
A:Gene: GDB:PTPRJ  
A:Cross-references: GDB:385040; OMIM:600925  
A:Map position: 19q13.4-19q13.4  
C:Function:  
A:Description: catalyzes hydrolysis of peptidyl-phosphotyrosine to peptidyl-tyrosine and  
C:Superfamily: protein-tyrosine-phosphatase, receptor type J; fibronectin type III repea  
C:Keywords: glycoprotein; phosphoprotein; phosphoric monoester hydrolase; receptor; tran  
F:1-35/Domain: signal sequence #status predicted <SIG>  
F:36-1337/Product: protein-tyrosine-phosphatase, receptor type J #status predicted <MAT>  
F:118-197/Domain: fibronectin type III repeat homology <3FNA>  
F:206-283/Domain: fibronectin type III repeat homology <3FNB>  
F:284-356/Domain: fibronectin type III repeat homology #status atypical <3FNC>  
F:365-445/Domain: fibronectin type III repeat homology <3FND>  
F:453-530/Domain: fibronectin type III repeat homology <3FNE>  
F:539-617/Domain: fibronectin type III repeat homology <3FNF>  
F:720-804/Domain: fibronectin type III repeat homology <3FNG>  
F:972-988/Domain: transmembrane #status predicted <TMN>  
F:1065-1287/Domain: protein-tyrosine-phosphatase homology <PTP>  
F:72,82,93,104,142,172,192,231,258,278,342,351,376,391,396,413,431,501,525,536,582,603,6

P:1239/Active site: Cys (phosphocysteine intermediate) #status predicted  
F:1245/Binding site: substrate phosphate (Arg) #status predicted

Query Match 35.5%; Score 239; DB 1; Length 1337;  
Best Local Similarity 38.1%; Pred. No. 4.2e-17;  
Matches 48; Conservative 22; Mismatches 50; Indels 6; Gaps 3;

QY 1 DFWGMMWNNCNIAMITREIEGGVKKCSYWPVSLKEPLEKHFHVLLENFOITQYFVI 60  
DB 1122 DFWRMVWERNVYAIIMLTCKVEQGRTKCEYWP--SKQADYGDITVAMTSEVLPEWTI 1179  
QY 61 RIFQIVKKSSTGKSHS-VKHLQFIKWPDHCTPASVDFFIKY---YVRKSHITGPLLHV 116  
DB 1180 RDTFVKNIQTSSEHPLRQFHFTSWPDHGVPDITDILLINFRYLVRDYMKQSPSPESILVHC 1239  
QY 117 TAGVGR 122  
DB 1240 SAGVGR 1245

RESULT 13  
A46151  
protein-tyrosine-phosphatase (EC 3.1.3.48) receptor type zeta precursor - human  
N:Alternate names: protein-tyrosine-phosphatase beta; protein-tyrosine-phosphatase xi  
C:Species: Homo sapiens (man)  
C:Date: 08-Dec-1993 #sequence\_revision 08-Feb-1996 #text\_change 22-Jun-1999  
C:Accession: A46151; A46700; B36065; S12054  
R:Krueger, N.X.; Saito, H.  
Proc. Natl. Acad. Sci. U.S.A. 89, 7417-7421, 1992  
A:Title: A human transmembrane protein-tyrosine-phosphatase, PTP zeta, is expressed i  
A:Reference number: A46151; MUID:92366472  
A:Accession: A46151  
A:Molecule type: mRNA  
A:Residues: 1-2314 <KRU>  
A:Cross-references: GB:M93426; NID:g190743; PIDN:AAA60225.1; PID:g190744  
A:Experimental source: fetal brain  
A:Note: sequence extracted from NCBI backbone (NCBIN:110851, NCBIP:110852)  
A:Note: sequence inconsistent with the nucleotide translation  
R:Levy, J.B.; Canoll, P.D.; Silvennoinen, O.; Barnea, G.; Morse, B.; Honegger, A.M.;  
nger, J.  
J. Biol. Chem. 268, 10573-10581, 1993  
A:Title: The cloning of a receptor-type protein tyrosine phosphatase expressed in the  
A:Reference number: A46700; MUID:93252948  
A:Accession: A46700  
A:Status: not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 1-1721,1729-2314 <LEV>  
A:Experimental source: brainstem  
A:Note: sequence extracted from NCBI backbone (NCBIP:131344)  
R:Kaplan, R.; Morse, B.; Huebner, K.; Croce, C.; Howk, R.; Ravera, M.; Ricca, G.; Jay  
Proc. Natl. Acad. Sci. U.S.A. 87, 7000-7004, 1990  
A:Title: Cloning of three human tyrosine phosphatases reveals a multigene family of r  
A:Reference number: A36065; MUID:90384936  
A:Accession: B36065  
A:Status: preliminary; not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 1749-1990,2047-2280 <KAP>  
A:Cross-references: GB:M34668; NID:g190738  
R:Krueger, N.X.; Streuli, M.; Saito, H.  
EMBO J. 9, 3241-3252, 1990  
A:Title: Structural diversity and evolution of human receptor-like protein tyrosine p  
A:Reference number: S12049; MUID:91006018  
A:Accession: S12054  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1479-2091 <KR2>  
A:Cross-references: GB:X54135; NID:g35795; PIDN:CAA38070.1; PID:g930104  
C:Genetics:  
A:Gene: GDB:PTPRZ1; PTPRZ; PTP2; HPTP2; PTP18; RPTPB  
A:Cross-references: GDB:127353; OMIM:176891  
A:Map position: 7q31.3-7q31.3  
C:Function:



Search completed: March 30, 2002, 08:16:25  
Job time: 3527 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 30, 2002, 07:10:48 ; Search time 77.75 Seconds  
(without alignments)  
116.231 Million cell updates/sec

Title: US-09-095-478A-8  
Perfect score: 673  
Sequence: 1 DFQGMWNNCNCVIAITRE.....VRKSHITGPLLHVHCTAGVGR 122

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_1101.\*  
1: /SIDS2/gcgdata/geneseq/geneseq/AA1980.DAT.\*  
2: /SIDS2/gcgdata/geneseq/geneseq/AA1981.DAT.\*  
3: /SIDS2/gcgdata/geneseq/geneseq/AA1982.DAT.\*  
4: /SIDS2/gcgdata/geneseq/geneseq/AA1983.DAT.\*  
5: /SIDS2/gcgdata/geneseq/geneseq/AA1984.DAT.\*  
6: /SIDS2/gcgdata/geneseq/geneseq/AA1985.DAT.\*  
7: /SIDS2/gcgdata/geneseq/geneseq/AA1986.DAT.\*  
8: /SIDS2/gcgdata/geneseq/geneseq/AA1987.DAT.\*  
9: /SIDS2/gcgdata/geneseq/geneseq/AA1988.DAT.\*  
10: /SIDS2/gcgdata/geneseq/geneseq/AA1989.DAT.\*  
11: /SIDS2/gcgdata/geneseq/geneseq/AA1990.DAT.\*  
12: /SIDS2/gcgdata/geneseq/geneseq/AA1991.DAT.\*  
13: /SIDS2/gcgdata/geneseq/geneseq/AA1992.DAT.\*  
14: /SIDS2/gcgdata/geneseq/geneseq/AA1993.DAT.\*  
15: /SIDS2/gcgdata/geneseq/geneseq/AA1994.DAT.\*  
16: /SIDS2/gcgdata/geneseq/geneseq/AA1995.DAT.\*  
17: /SIDS2/gcgdata/geneseq/geneseq/AA1996.DAT.\*  
18: /SIDS2/gcgdata/geneseq/geneseq/AA1997.DAT.\*  
19: /SIDS2/gcgdata/geneseq/geneseq/AA1998.DAT.\*  
20: /SIDS2/gcgdata/geneseq/geneseq/AA1999.DAT.\*  
21: /SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT.\*  
22: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	673	100.0	122	AAW89252	Rat PTP10.
2	566	84.1	405	AAW89251	Mouse PTP05 isofo
3	566	84.1	426	AAW89249	Mouse PTP05. Mus
4	566	84.1	463	AAW89250	Mouse PTP05 isofo
5	268.5	39.9	2466	AAW71498	Human protein tyro
6	268.5	39.9	2466	AAW75999	Intracellular prot
7	268.5	39.9	2466	AAW90272	Human PTP1 phosph
8	268.5	39.9	2485	AAW19343	Amino acid sequen
9	250	37.1	913	AAW12522	Protein tyrosine p
10	250	37.1	913	AAW25156	Human PTP1H protei
11	240	35.7	240	AAW20745	Human R-PTPase bet

12	240	35.7	242	22	AAW59384	Human protein tyro
13	240	35.7	579	21	AAW19773	Mouse vascular-end
14	239.5	35.6	253	22	AAW59374	Murine protein tyr
15	239	35.5	1337	16	AAW85203	huDEP-1. Homo sap
16	238	35.4	2308	15	AAW57902	Human RPTP-beta.
17	238	35.4	2308	22	AAW59586	Human receptor-tyr
18	237	35.2	1911	16	AAW71726	Human PTP-OB. Hom
19	237	35.2	1911	18	AAW27225	Human protein tyro
20	237	35.2	1911	20	AAW94027	Human protein tyro
21	237	35.2	1911	22	AAU01459	Human protein tyro
22	236.5	35.1	623	22	AAW25675	Human protein sequ
23	232.5	34.5	1499	22	AAW25768	Human protein sequ
24	231	34.3	1501	16	AAW72858	Rat receptor type-
25	231	34.3	1997	21	AAW19774	Human protein tyro
26	230.5	34.2	1430	19	AAW49907	Human pancreatic c
27	230	34.2	442	21	AAW56372	Human prostate can
28	230	34.2	607	21	AAW81783	Human protein tyro
29	230	34.2	607	21	AAW56098	LAR tyrosine phosph
30	230	34.2	647	22	AAW23746	Human EST encoded
31	230	34.2	647	22	AAU14379	Human novel protei
32	230	34.2	1897	21	AAW19712	Human protein tyro
33	230	34.2	1897	21	AAW81785	Human protein tyro
34	230	34.2	1897	21	AAW56100	LAR tyrosine phosph
35	230	34.2	1907	22	AAU14143	Human novel protei
36	226	33.6	253	22	AAW59383	Human protein tyro
37	226	33.6	775	21	AAW67250	Mouse protein tyro
38	225	33.4	260	22	AAW59367	Human protein tyro
39	225	33.4	260	22	AAW59388	Human protein tyro
40	224.5	33.4	415	12	AAW14114	Non-receptor linke
41	224.5	33.4	1436	19	AAW41361	Receptor protein t
42	224	33.3	245	22	AAW59389	Human protein tyro
43	224	33.3	1149	22	AAW41048	Human polypeptide
44	221.5	32.9	249	22	AAW59371	Human protein tyro
45	221	32.8	780	21	AAW67252	Human protein tyro

ALIGNMENTS

RESULT 1  
AAW89252  
ID AAW89252 standard; Protein; 122 AA.

XX AAW89252;

AC AAW89252;

XX j0-MAR-1999 (first entry)

DT j0-MAR-1999 (first entry)

XX Rat PTP10.

XX Rat PTP10.

XX PTP04; PTP05; PTP10; SAD; ALP; ALK-7; protein tyrosine phosphatase;

XX type I receptor serine/threonine kinase; cancer; leukaemia; lymphoma;

XX neurodegenerative disease; neuronal survival; Alzheimer's disease;

XX Parkinson's disease; Huntington's disease.

XX Rattus sp.

XX WO98493317-A2.

XX 05-NOV-1998.

XX 27-APR-1998; 98WO-US08439.

XX 23-OCT-1997; 97US-0063595.

XX 28-APR-1997; 97US-0044428.

XX 20-MAY-1997; 97US-0047222.

XX 11-JUN-1997; 97US-0049477.

XX 11-JUN-1997; 97US-0049756.

XX 18-JUN-1997; 97US-0049914.

XX (SUGEN-) SUGEN INC.

XX App H. Clary D. Courtneidge SA. Hui TH. Jallal B.

XX Markby D. Onrust S. Peles E. Plowman GD.

```

XX WPI: 1999-009434/01.
DR N-PSDB; AAV811747.
XX
XX New nucleic acid encoding specific protein tyrosine phosphatases -
PT useful for identifying specific modulators for treatment and
PT prevention of cancer and neurodegenerative disease
XX
XX Claim 2; Page 160; 193pp; English.
XX
XX The present invention describes isolated, enriched or purified nucleic
CC acids encoding PTP04, SAD, PTP05, PTP10, ALP and ALK-7 proteins. The
CC present sequence represents rat PTP10. The above proteins, other than
CC ALK-7, are protein tyrosine phosphatases (PTPs) and are used to identify
CC substances that modulate their activity (i.e. agonists and antagonists,
CC including NBP) in vivo or in vitro. These substances are used to treat
CC or prevent diseases associated with abnormal signal transduction
CC pathways that involve the proteins, particularly cancer (e.g. leukaemia
CC and lymphoma), while modulators of ALK-7 (which is a type I receptor
CC serine/threonine kinase) are used to promote neuronal survival,
CC particularly for treating Alzheimer's, Parkinson's or Huntington's
CC diseases. Nucleic acid fragments of the polynucleotides encoding the
CC proteins can be used as probes to identify and clone related sequences;
CC to detect protein-encoded RNA; to generate transgenic animals and in
CC gene therapy (optionally after mutation). Ab are used to determine the
CC proteins.
XX
XX Sequence 122 AA;
SQ
Query Match 100.0%; Score 673; DB 20; Length 122;
Best Local Similarity 100.0%; Pred. No. 9.1e-79; Indels 0; Gaps 0;
Matches 122; Conservative 0; Mismatches 0;
QY 1 DFMGMWENNVCNVIAMITREIEGGVVKCCSYWPSVSLKEPLEKHFHVLLENFOITQYFVI 60
DB 1 dfwgmmwennvcnviamitreieggvkkccsywpsvslkeplekfhfvllefnfqtqyfv 60
QY 61 RFIQIVKSTGKSHVYKHLQFIKWPDPHTPASVDFFIKYVRYVRKSHITGPLLHVHCTAGV 120
DB 61 rfiqivkstgkshvkhlfqikwpdhtpasvdffikyrvyvrkshitgpllvhctagv 120
QY 121 GR 122
DB 121 gr 122
RESULT 2
ID AAW89251
AC AAW89251;
XX
XX 10-MAR-1999 (first entry)
XX Mouse PTP05 isoform #2.
XX
XX PTP04; PTP05; PTP10; SAD; ALP; ALK-7; protein tyrosine phosphatase;
KW type I receptor serine/threonine kinase; cancer; leukaemia; lymphoma;
KW neurodegenerative disease; neuronal survival; Alzheimer's disease;
KW Parkinson's disease; Huntington's disease.
XX
XX Mus sp.
XX
XX WO9849317-A2.
XX
XX 05-NOV-1998.
XX
XX 27-APR-1998; 98WO-US08439.
XX
XX 23-OCT-1997; 97US-0063595.
XX 28-APR-1997; 97US-0044428.
XX 20-MAY-1997; 97US-0047222.
XX

```

---

```

PR 11-JUN-1997; 97US-0049477.
PR 11-JUN-1997; 97US-0049756.
PR 18-JUN-1997; 97US-0049914.
XX
XX (SUGE-) SUGEN INC.
XX
XX App H, Clary D, Courtneidge SA, Hui TH, Jallal B;
PI Markby D, Onrust S, Peles E, Plowman GD;
XX
XX WPI: 1999-009434/01.
DR N-PSDB; AAV811746.
XX
XX New nucleic acid encoding specific protein tyrosine phosphatases -
PT useful for identifying specific modulators for treatment and
PT prevention of cancer and neurodegenerative disease
XX
XX Claim 2; Page 158-160; 193pp; English.
XX
XX The present invention describes isolated, enriched or purified nucleic
CC acids encoding PTP04, SAD, PTP05, PTP10, ALP and ALK-7 proteins. The
CC present sequence represents mouse PTP05. The above proteins, other than
CC ALK-7, are protein tyrosine phosphatases (PTPs) and are used to identify
CC substances that modulate their activity (i.e. agonists and antagonists,
CC including NBP) in vivo or in vitro. These substances are used to treat
CC or prevent diseases associated with abnormal signal transduction
CC pathways that involve the proteins, particularly cancer (e.g. leukaemia
CC and lymphoma), while modulators of ALK-7 (which is a type I receptor
CC serine/threonine kinase) are used to promote neuronal survival,
CC particularly for treating Alzheimer's, Parkinson's or Huntington's
CC diseases. Nucleic acid fragments of the polynucleotides encoding the
CC proteins can be used as probes to identify and clone related sequences;
CC to detect protein-encoded RNA; to generate transgenic animals and in
CC gene therapy (optionally after mutation). Ab are used to determine the
CC proteins.
XX
XX Sequence 405 AA;
SQ
Query Match 84.1%; Score 566; DB 20; Length 405;
Best Local Similarity 83.6%; Pred. No. 2.5e-64;
Matches 102; Conservative 7; Mismatches 13; Indels 0; Gaps 0;
QY 1 DFMGMWENNVCNVIAMITREIEGGVVKCCSYWPSVSLKEPLEKHFHVLLENFOITQYFVI 60
DB 244 dfwgmmwennvcnviamitreieggvkkccsywpsvslkeplekfhfvllefnfvtqyfv 303
QY 61 RFIQIVKSTGKSHVYKHLQFIKWPDPHTPASVDFFIKYVRYVRKSHITGPLLHVHCTAGV 120
DB 304 rfiqivkstgkshvkhlfqikwpdhtpasvdffikyrvyvrkshitgpllvhctagv 363
QY 121 GR 122
DB 364 gr 365
RESULT 3
AAW89249;
ID AAW89249 standard; Protein; 426 AA.
XX
XX AAW89249;
XX
XX 10-MAR-1999 (first entry)
XX
XX Mouse PTP05.
XX
XX PTP04; PTP05; PTP10; SAD; ALP; ALK-7; protein tyrosine phosphatase;
KW type I receptor serine/threonine kinase; cancer; leukaemia; lymphoma;
KW neurodegenerative disease; neuronal survival; Alzheimer's disease;
KW Parkinson's disease; Huntington's disease.
XX
XX Mus sp.
XX
XX WO9849317-A2.
XX

```

```

XX 05-NOV-1998.
XX 27-APR-1998; 98WO-US08439.
XX 23-OCT-1997; 97US-0063595.
XX 28-APR-1997; 97US-0044428.
XX 20-MAY-1997; 97US-0047222.
XX 11-JUN-1997; 97US-0049477.
XX 11-JUN-1997; 97US-0049756.
XX 18-JUN-1997; 97US-0049914.
XX (SUGEN-) SUGEN INC.
XX App H, Clary D, Courtneidge SA, Hui TH, Jallal B;
XX Markby D, Onrust S, Peles E, Plowman GD;
XX WPI; 1999-009434/01.
XX N-PSDB; AAV81744.
XX New nucleic acid encoding specific protein tyrosine phosphatases -
XX useful for identifying specific modulators for treatment and
XX prevention of cancer and neurodegenerative disease
XX Claim 2; Page 155-157; 193pp; English.
XX The present invention describes isolated, enriched or purified nucleic
XX acids encoding PTP04, SAD, PTP05, PTP10, ALP and ALK-7 proteins. The
XX present sequence represents mouse PTP05. The above proteins, other than
XX ALK-7, are protein tyrosine phosphatases (PTPs) and are used to identify
XX substances that modulate their activity (i.e. agonists and antagonists,
XX including NBP) in vivo or in vitro. These substances are used to treat
XX or prevent diseases associated with abnormal signal transduction
XX pathways that involve the proteins, particularly cancer (e.g. leukaemia
XX and lymphoma), while modulators of ALK-7 (which is a type I receptor
XX serine/threonine kinase) are used to promote neuronal survival,
XX particularly for treating Alzheimer's, Parkinson's or Huntington's
XX diseases. Nucleic acid fragments of the polynucleotides encoding the
XX proteins can be used as probes to identify and clone related sequences;
XX to detect protein-encoded RNA; to generate transgenic animals and in
XX gene therapy (optionally after mutation). Ab are used to determine the
XX proteins.
XX Sequence 426 AA;
XX
XX Query Match 84.1%; Score 566; DB 20; Length 426;
XX Best Local Similarity 83.6%; Pred. No. 2.6e-64;
XX Matches 102; Conservative 7; Mismatches 13; Indels 0; Gaps 0;
XX
XX QY 1 DFWGMWNNCNCVNIAMITREIEGGVKKCSYWPVSLKEPKEFHVLLENFOITQYFYVI 60
XX Db 244 dfwqmvlenncncvniamitreiecgvikcsywpvslkeplefshvlfethvftqyftv 303
XX
XX QY 61 RFQIVKSTGKSHSVKHLQFTKPDHGTGPASVDFFIKYVYVRKSHITGPLLHVHCTAGV 120
XX Db 304 rfvqivkstksgkscvkhqlfctkwpdhtgtpasadffikyvrvrkshitgpllvhcsagv 363
XX
XX QY 121 GR 122
XX Db 364 gr 365
XX
XX RESULT 4
XX ID AAW89250 standard; Protein; 463 AA.
XX AC AAW89250;
XX DT 10-MAR-1999 (first entry)
XX DE Mouse PTP05 isoform #1.
XX

```

```

KW PTP04; PTP05; SAD; ALP; ALK-7; protein tyrosine phosphatase;
KW type I receptor serine/threonine kinase; cancer; leukaemia; lymphoma;
KW neurodegenerative disease; neuronal survival; Alzheimer's disease;
KW parkinson's disease; Huntington's disease.
XX
XX OS Mus sp.
XX PN WO9849317-A2.
XX PD 05-NOV-1998.
XX XX
XX PF 27-APR-1998; 98WO-US08439.
XX PR 23-OCT-1997; 97US-0063595.
XX PR 28-APR-1997; 97US-0044428.
XX PR 20-MAY-1997; 97US-0047222.
XX PR 11-JUN-1997; 97US-0049477.
XX PR 11-JUN-1997; 97US-0049756.
XX PR 18-JUN-1997; 97US-0049914.
XX
XX (SUGEN-) SUGEN INC.
XX App H, Clary D, Courtneidge SA, Hui TH, Jallal B;
XX Markby D, Onrust S, Peles E, Plowman GD;
XX WPI; 1999-009434/01.
XX N-PSDB; AAV81745.
XX New nucleic acid encoding specific protein tyrosine phosphatases -
XX useful for identifying specific modulators for treatment and
XX prevention of cancer and neurodegenerative disease
XX Claim 2; Page 157-158; 193pp; English.
XX The present invention describes isolated, enriched or purified nucleic
XX acids encoding PTP04, SAD, PTP05, PTP10, ALP and ALK-7 proteins. The
XX present sequence represents mouse PTP05. The above proteins, other than
XX ALK-7, are protein tyrosine phosphatases (PTPs) and are used to identify
XX substances that modulate their activity (i.e. agonists and antagonists,
XX including NBP) in vivo or in vitro. These substances are used to treat
XX or prevent diseases associated with abnormal signal transduction
XX pathways that involve the proteins, particularly cancer (e.g. leukaemia
XX and lymphoma), while modulators of ALK-7 (which is a type I receptor
XX serine/threonine kinase) are used to promote neuronal survival,
XX particularly for treating Alzheimer's, Parkinson's or Huntington's
XX diseases. Nucleic acid fragments of the polynucleotides encoding the
XX proteins can be used as probes to identify and clone related sequences;
XX to detect protein-encoded RNA; to generate transgenic animals and in
XX gene therapy (optionally after mutation). Ab are used to determine the
XX proteins.
XX Sequence 463 AA;
XX
XX Query Match 84.1%; Score 566; DB 20; Length 463;
XX Best Local Similarity 83.6%; Pred. No. 2.9e-64;
XX Matches 102; Conservative 7; Mismatches 13; Indels 0; Gaps 0;
XX
XX QY 1 DFWGMWNNCNCVNIAMITREIEGGVKKCSYWPVSLKEPKEFHVLLENFOITQYFYVI 60
XX Db 281 dfwqmvlenncncvniamitreiecgvikcsywpvslkeplefshvlfethvftqyftv 340
XX
XX QY 61 RFQIVKSTGKSHSVKHLQFTKPDHGTGPASVDFFIKYVYVRKSHITGPLLHVHCTAGV 120
XX Db 341 rfvqivkstksgkscvkhqlfctkwpdhtgtpasadffikyvrvrkshitgpllvhcsagv 400
XX
XX QY 121 GR 122
XX Db 401 gr 402
XX
XX RESULT 5
XX ID AAW71498

```







AC AAW12522;  
 XX 22-MAY-1997 (first entry)  
 XX Protein tyrosine phosphatase that localises to focal adhesion.  
 DE  
 XX  
 XX Protein tyrosine phosphatase; PTPH1; focal adhesion;  
 KW protein tyrosine kinase; malignancy; cancer; gene therapy;  
 KW retrovirus; vector.  
 XX  
 XX Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT Region 30..357  
 FT /note= "N-terminal region shows homology with the  
 FT N-terminal regions of the talin family"  
 FT Region 357..663  
 FT /note= "central region includes sequences with  
 FT features of sites of phosphorylation by  
 FT casein kinase II and p34cdc2"  
 FT Modified-site 372  
 FT /label= Phosphorylation  
 FT /note= "putative p34cdc2 phosphorylation site"  
 FT Modified-site 381  
 FT /label= Phosphorylation  
 FT /note= "putative p34cdc2 phosphorylation site"  
 FT Modified-site 424..428  
 FT /label= Phosphorylation  
 FT /note= "putative casein kinase II phosphorylation  
 FT site"  
 FT Modified-site 438..442  
 FT /label= Phosphorylation  
 FT /note= "putative casein kinase II phosphorylation  
 FT site"  
 FT Modified-site 489..492  
 FT /label= Phosphorylation  
 FT /note= "putative casein kinase II phosphorylation  
 FT site"  
 FT Modified-site 514..517  
 FT /label= Phosphorylation  
 FT /note= "putative casein kinase II phosphorylation  
 FT site"  
 FT Modified-site 543..547  
 FT /label= Phosphorylation  
 FT /note= "putative casein kinase II phosphorylation  
 FT site"  
 FT Modified-site 607..610  
 FT /label= Phosphorylation  
 FT /note= "putative casein kinase II phosphorylation  
 FT site"  
 FT Region 664..931  
 FT /note= "C-terminal region shows homology to "  
 FT  
 XX US5595911-A.  
 XX 21-JAN-1997.  
 XX 14-MAR-1990; 9005-0494036.  
 XX 01-MAR-1991; 9105-0663579.  
 XX 14-MAR-1990; 9005-0494036.  
 XX 16-AUG-1993; 9305-0107420.  
 XX (COLD-) COLD SPRING HARBOR LAB.  
 XX  
 XX Tonks NK;  
 XX  
 XX WPI; 1997-107583/10.  
 DR N-PSDB; AAT58627.  
 DR  
 XX DNA encoding protein tyrosine phosphatase - for gene therapy of  
 PT cancer  
 XX

PS Disclosure; Fig 1A-B; 12pp; English.  
 XX  
 CC A protein tyrosine phosphatase (PTPH1) (AAW12522) catalyses the  
 CC dephosphorylation of proteins in which tyrosyl residues have been  
 CC phosphorylated through the action of a protein tyrosine kinase  
 CC (PTK). It localises to focal adhesions, a major site of action of  
 CC oncogenic PTKs. PTPH1 is the product of a cDNA clone (AAT58627)  
 CC cbtd. from HeLa cells. The PTPH1 cDNA can be incorporated into a  
 CC vector (pref. retroviral) to allow expression of PTPH1 in mammalian  
 CC in sufficient quantities to overcome or counteract PTK activity.  
 CC phosphorylation of tyrosine residues at abnormal levels is  
 CC prevented or reversed, resulting in the prevention or reversal of  
 CC malignancy of cells.  
 XX  
 XX Sequence 913 AA;  
 SQ  
 Query Match 37.1%; Score 250; DB 18; Length 913;  
 Best Local Similarity 37.9%; Pred. No. 3.3e-23;  
 Matches 47; Conservative 23; Mismatches 48; Indels 6; Gaps 3;  
 QY 2 FWMGMWNNCNVIAITREIEGGVVKCCSVYPSVSLKKEPLEFKH--FHVLLNFOITQYEV 59  
 DB 728 fqvvdqklsilvmltlttergrtkchqwp---dpdvnmhggfhiqcqsedctiayv 784  
 QY 60 IRIFQIVYKSTGKSHSVKHLQFIKWPDPGTPASVDFFIKYRYVRKSHI-TGPLLVHCTA 118  
 DB 785 sremlvntqtgeehvtthlyvawpdhgiiddssdflefvyrvslrvdseplvhcsa 844  
 QY 1-9 GVGR 122  
 DB .845 g1gr 848  
 RESULT 10  
 AAY25156  
 ID AAY25156 standard; Protein; 913 AA.  
 XX  
 AC AAY25156;  
 XX  
 DT 07-SEP-1999 (first entry)  
 XX  
 XX Human PTPH1 protein.  
 XX  
 KW PTPH1; human; protein tyrosine phosphatase; focal adhesion; cancer;  
 KW localisation; treatment; overexpression; oncogenic; cell transformation;  
 KW prevention; phosphotyrosine; disease; malignant.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key' Location/Qualifiers  
 FT Region 30..357  
 FT /note= "region of homology to the N-terminal domain  
 FT of band 4.1, ezrin and talin. This region is  
 FT known to be important for localisation to focal  
 FT adhesions"  
 FT Modified-site 372  
 FT /note= "potential p34cdc2 phosphorylation site"  
 FT Modified-site 381  
 FT /note= "potential p34cdc2 phosphorylation site"  
 FT Modified-site 424..428  
 FT /note= "potential casein kinase II phosphorylation site"  
 FT Modified-site 434  
 FT /note= "potential p34cdc2 phosphorylation site"  
 FT Modified-site 438..442  
 FT /note= "potential casein kinase II phosphorylation site"  
 FT Modified-site 489..492  
 FT /note= "potential casein kinase II phosphorylation site"  
 FT Modified-site 514..518  
 FT /note= "potential casein kinase II phosphorylation site"  
 FT Modified-site 543..547  
 FT /note= "potential casein kinase II phosphorylation site"  
 FT Modified-site 607..610

FT /note- "potential casein kinase II phosphorylation site"  
 XX US863781-A.  
 PN  
 XX  
 XX  
 PD 26-JAN-1999.  
 XX  
 XX  
 PF 04-DEC-1996; 96US-0759536.  
 XX  
 XX 01-MAR-1991; 91US-0663579.  
 PR 14-MAR-1990; 90US-0494036.  
 PR 16-AUG-1993; 93US-0107420.  
 PR 04-DEC-1996; 96US-0759536.  
 XX  
 XX (COLD-) COLD SPRING HARBOR LAB.  
 PA  
 XX  
 PI Tonks NK;  
 DR WPI; 1999-131308/11.  
 DR N-PSDB; AAX78463.  
 XX  
 XX Protein tyrosine phosphatase PTPH1 - encoded by DNA of HeLa cells  
 PT  
 PS Claim 1; Fig 1A-B; 12pp; English.  
 XX This sequence represents a novel protein tyrosine phosphatase, PTPH1,  
 CC isolated from HeLa cells. The protein of the invention appears to  
 CC localise to focal adhesions and is therefore potentially useful in the  
 CC treatment of cancer. Overexpression of PTPH1 can be used to counter the  
 CC effects of oncogenic protein tyrosine kinases such as those of  
 CC transforming viruses and for interfering with or reversing cell  
 CC transformation. This would provide a means of preventing or reversing  
 CC abnormally high levels of phosphotyrosine associated with any disease or  
 CC condition such as preventing or reversing malignancy associated with the  
 CC activity of a protein tyrosine kinase.  
 XX  
 SQ Sequence 913 AA;

Query Match 37.1%; Score 250; DB 20; Length 913;  
 Best Local Similarity 37.9%; Pred. No. 3.3e-23;  
 Matches 47; Conservative 23; Mismatches 48; Indels 6; Gaps 3;  
 QY 2 FWGMWENNVCNVIAMITREIEGGVTKCCSYWPSVLKEPLEFKH--FHVLLLENFQITQYFV 59  
 DB 728 fqvvdqklsilvmtltltergrtkchqwp---dppdmhghgfhqcgdsedctiayv 784  
 QY 60 IRIFOIVKVKSTGKSHVHLQFIKWPDHGTPASVDFFIKYVYVRKSHI-TGPLLVHCTA 118  
 DB 785 sremivntqtgeetvthlgyavpdhgipddssdflefnyvrsrlrdsepvlvhcsa 844  
 QY 119 GVGR 122  
 DB 845 gigr 848  
 RESULT 11  
 AAR20745  
 ID AAR20745 standard; Protein; 240 AA.  
 XX  
 AC AAR20745;  
 XX  
 DT 28-MAY-1992 (first entry)  
 XX  
 DE Human R-PTPase beta first conserved phosphatase.  
 XX  
 KW Receptor-type protein tyrosine phosphatase; cellular metabolism;  
 KW cancer; diabetes.  
 XX  
 OS Homo sapiens.  
 XX  
 PN W09201050-A.  
 XX  
 PD 23-JAN-1992.

XX 11-JUL-1991; 91WO-US04892.  
 PF  
 XX  
 XX 26-FEB-1991; 91US-0654188.  
 PR 11-JUL-1990; 90US-0551270.  
 XX  
 XX (U'NY-) NEW YORK UNIV.  
 PA  
 XX  
 XX Schlessinger J;  
 XX  
 DR WPI; 1992-056865/07.  
 XX  
 XX Human receptor-type protein tyrosine phosphatase - has DNA  
 PT encoding it and antibodies specific for it, useful for screening  
 PF drugs affecting R-PTPase activity, and detection of mutant genes  
 XX  
 XX Claim 5; Fig 5A; 77pp; English.  
 XX  
 XX This amino acid sequence is that of human receptor-type protein  
 CC tyrosine phosphatase (R-PTPase) beta first conserved phosphatase. It  
 CC is useful in methods for screening drugs and other agents which are  
 CC capable of activating or inhibiting the R-PTPase activity and thereby  
 CC affecting major pathways of cellular metabolism. Activation of  
 CC R-PTPases, leading to dephosphorylation would serve as a counter-  
 CC regulatory mechanism to prevent or inhibit growth, and may serve as  
 CC an endogenous regulatory mechanism against cancer. Mutation or  
 CC dysregulation of this receptor/enzyme system may promote susceptibility  
 CC to cancer, diabetes, or other diseases associated with alterations in  
 CC cellular phosphotyrosine metabolism. It can be used to raise antibodies  
 CC which can be used in immunoassays to determine the presence and amt.  
 CC of R-PTPases, or in immunoelectron microscopy for in situ detection of  
 CC R-PTPase. See also AAR20743-R20748.  
 XX  
 SQ Sequence 240 AA;  
 Query Match 35.7%; Score 240; DB 13; Length 240;  
 Best Local Similarity 37.6%; Pred. No. 1.2e-22;  
 Matches 50; Conservative 24; Mismatches 43; Indels 16; Gaps 5;  
 QY 1 DFWMWENNVCNVIAMITREIEGGVTKCCSYWPSVLKEPLEFKHVLLENFQITQYFVI 60  
 DB 61 dfwrmiwheveivmitnlvkegrkrkdqywpdgse--eygnflvtksqvlavtvcv 118  
 QY 61 RIFQI----VKKSTGK--SHSVKHLQFIKWPDHGTPASVDFFIKYVYVRKS-----HIT 109  
 DB 119 rntitratkikkgkqrpgrvvtqhytqwpdhgvp---nynlpvltfvrkaayakrhav 175  
 QY 110 GPLLVHCTAGVGR 122  
 DB 176 gpvvvhcsagvgr 188  
 RESULT 12  
 AAB59384  
 ID AAB59384 standard; Protein; 242 AA.  
 XX  
 AC AAB59384;  
 XX  
 DT 21-MAR-2001 (first entry)  
 XX  
 DE Human protein tyrosine phosphatase #15.  
 XX  
 KW Protein tyrosine phosphatase; human; mouse; fruit fly; PTP;  
 KW substrate trapping.  
 XX  
 OS Homo sapiens.  
 XX  
 PN W0200075339-A1.  
 XX  
 PD 14-DEC-2000.  
 XX  
 PF 24-MAY-2000; 2000WO-US14211.



PA (COLD-) COLD SPRING HARBOR LAB.  
XX  
PI Tonks NK, Zhang S;  
XX  
DR WPI; 2001-080598/09.  
XX  
PT New substrate trapping mutant protein tyrosine phosphatases (PTP) in  
PT which the wild type PTP catalytic domain invariant aspartate is  
PT replaced with an unphosphorylated amino acid, useful in gene therapy  
XX  
PS Disclosure: Fig 1; 109pp; English.  
XX  
CC The present invention provides substrate trapping mutant protein tyrosine  
CC phosphatases (PTPs). They can be used to reduce the activity of tyrosine  
CC phosphorylated proteins and to screen for modulators capable of altering  
CC the binding of protein tyrosine phosphatases to their substrate. These  
CC may be used in disease diagnosis and treatment.  
XX  
SQ Sequence 253 AA;  
XX  
Query Match 35.6%; Score 239.5; DB 22; Length 253;  
Best Local Similarity 38.4%; Pred. No. 1.4e-22;  
Matches 48; Conservative 20; Mismatches 52; Indels 5; Gaps 3;  
QY 1 DFWGMWNNCNCVIAITREIEGGVKKCSYWPVSLKEPLEKHFHVLLENFOITQYFV 59  
DB 70 dfwrmlweqkatvnmvtceegnrnkcaeywpsmeegtrafkdivvtindhkrcpdyii 129  
QY 60 RIFQIVKSTGKSHVHLQFIKWPDHGTPASVDFFIKYVR--KSHITGPLLHVHCT 117  
DB 130 Invankkekatgr--evthiqftswpdhgvpedphllklrrvnaafsnffsgpiyvhes 187  
QY 118 AGVGR 122  
DB 188 agvgr 192  
RESULT 15  
AAR85203  
ID AAR85203 standard; Protein; 1337 AA.  
XX  
AC AAR85203;  
XX  
DT 12-FEB-1996 (first entry)  
XX  
DE huDEP-1.  
XX  
KW Density enhanced Type III receptor-like protein tyrosine phosphatase;  
KW huDEP-1.  
XX  
OS Homo sapiens.  
XX  
PN W09530008-A1.  
XX  
PD 09-NOV-1995.  
XX  
PF 03-MAY-1995; 95WO-US05512.  
XX  
PR 03-MAY-1994; 94US-0237940.  
XX  
PA (COLD-) COLD SPRING HARBOR LAB.  
XX  
PI Oestman A, Tonks NK;  
XX  
DR WPI; 1995-393079/50.  
DR N-PSDB; AAT06027.  
XX  
PT New density enhanced protein tyrosine phosphatase - used to develop  
PT prods. to modify transcription, translation and/or activity of  
PT tyrosine phosphatase(s).  
XX  
PS Claim 4; Page 34-38; 51pp; English.

XX A cDNA clone was obtd. (see AAT06027) from a HeLa cell cDNA library  
CC that encoded a novel density-enhanced type III receptor-like PTP,  
CC designated huDEP-1 (AAR85203). huDEP-1 is useful for the study of PTPs  
CC and for the development of therapeutic or prophylactic cpds. e.g. for  
CC prevention of abnormal or malignant cell growth.  
XX  
SQ Sequence 1337 AA;  
XX  
Query Match 35.5%; Score 239; DB 16; Length 1337;  
Best Local Similarity 38.1%; Pred. No. 1.4e-21;  
Matches 48; Conservative 22; Mismatches 50; Indels 6; Gaps 3;  
QY 1 DFWGMWNNCNCVIAITREIEGGVKKCSYWPVSLKEPLEKHFHVLLENFOITQYFVI 60  
DB 1122 dfwrmwvweknvyailmltkcvegrtkceeywp--skqadqdygditvamtseivlpewtl 1179  
QY 61 RIFQIVKSTGKSHVHLQFIKWPDHGTPASVDFFIKY---VR-YVRKSHITGPLLHVHC 116  
DB 180 rdtvknigtsseshplrqfhtswpdhgvdpdtdllinfrylvrdymkgsppespllvhc 1239  
QY 117 TAGVGR 122  
DB 1240 sagvgr 1245  
Search completed: March 30, 2002, 08:13:45  
Job time: 3777 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2002, 05:16:17 ; Search time 1554.02 Seconds  
(without alignments)  
2212.744 Million cell updates/sec

Title: US-09-095-478A-4

Perfect score: 320

Sequence: 1 GAAATAATGTAATGTTAT.....ATTACAGACCCCTCCTTGT 320

Scoring table:

Gapop 60.0 , Gapext 60.0

Searched: 11351937 seqs, 5372889281 residues

Word size : 0

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST:\*  
1: em\_estfun:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estom:\*  
5: em\_estpl:\*  
6: em\_estba:\*  
7: em\_estro:\*  
8: em\_estov:\*  
9: em\_hic:\*  
10: gb\_estl:\*  
11: gb\_est2:\*  
12: gb\_hic:\*  
13: gb\_gss:\*  
14: em\_gss\_fun:\*  
15: em\_gss\_hum:\*  
16: em\_gss\_inv:\*  
17: em\_gss\_pln:\*  
18: em\_gss\_pro:\*  
19: em\_gss\_rod:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	50	15.6	1842	12 AK017003	AK017003 Mus muscu
2	27	8.4	934	13 AQ744125	AQ744125 HS_5507_A
3	26	8.1	544	10 AF247150	AF247150 AF247150
4	20	6.2	406	10 AA633769	AA633769 ac27d11.s
5	20	6.2	550	10 AA129435	AA129435 zn85f02.s
6	20	6.2	849	13 AZ681898	AZ681898 ENTLJ27TR
7	20	6.2	850	13 AZ682025	AZ682025 ENTLR91TF
8	20	6.2	893	13 AZ689826	AZ689826 ENTLT42TF
9	19	5.9	259	10 BE711384	BE711384 PM2-H7068
10	19	5.9	615	10 AW559472	AW559472 EST314520
11	19	5.9	666	13 AZ083136	AZ083136 RPCI-23-3
12	18	5.6	136	10 AA230263	AA230263 nc13a04.f

C 13	18	5.6	155	11	BF662254	BF662254 maa80e02.
C 14	18	5.6	164	10	AI784154	AI784154 th88d03.x
C 15	18	5.6	177	10	AV298638	AV298638 AV298638
C 16	18	5.6	190	10	BB439182	BB439182 BB439182
C 17	18	5.6	192	11	D45598	D45598 HUMGS02790
C 18	18	5.6	199	10	AW072956	AW072956 xa43c01.x
C 19	18	5.6	219	10	AI864084	AI864084 wg63c11.x
C 20	18	5.6	228	10	BB214089	BB214089 BB214089
C 21	18	5.6	235	10	AV320214	AV320214 AV320214
C 22	18	5.6	243	10	AI991947	AI991947 ws42f11.x
C 23	18	5.6	245	10	AV320933	AV320933 AV320933
C 24	18	5.6	252	10	AV326992	AV326992 AV326992
C 25	18	5.6	253	11	BC670484	BC670484 DRNBGG04
C 26	18	5.6	259	10	AA558169	AA558169 n127c11.s
C 27	18	5.6	272	10	AA884408	AA884408 am16a04.s
C 28	18	5.6	275	10	AA047439	AA047439 zk73b02.s
C 29	18	5.6	276	10	BB052509	BB052509 BB052509
C 30	18	5.6	279	10	BB045084	BB045084 BB045084
C 31	18	5.6	279	10	BB303287	BB303287 BB303287
C 32	18	5.6	279	10	BB490959	BB490959 BB490959
C 33	18	5.6	279	10	BB565548	BB565548 BB565548
C 34	18	5.6	281	10	BB386000	BB386000 BB386000
C 35	18	5.6	281	11	R07867	R07867 yf16d05.s1
C 36	18	5.6	284	10	BB345232	BB345232 BB345232
C 37	18	5.6	284	10	BB435655	BB435655 BB435655
C 38	18	5.6	285	10	BB092661	BB092661 BB092661
C 39	18	5.6	286	10	BB379014	BB379014 BB379014
C 40	18	5.6	289	10	BB316043	BB316043 BB316043
C 41	18	5.6	291	10	BB303480	BB303480 BB303480
C 42	18	5.6	296	11	F37150	F37150 HSPD35253.H
C 43	18	5.6	298	11	BF466708	BF466708 UI-M-CG0p
C 44	18	5.6	299	10	AA508465	AA508465 nh66c02.s
C 45	18	5.6	300	10	AV127357	AV127357 AV127357

#### ALIGNMENTS

RESULT 1

AK017003 1842 bp mRNA 05-JUL-2001  
LOCUS Mus musculus adult male testis cDNA, RIKEN full-length enriched library, clone:4933431F18, full insert sequence.  
DEFINITION  
ACCESSION AK017003  
VERSION AK017003.1 GI:12856041  
KEYWORDS CAP trapper.  
SOURCE Mus musculus (strain:C57BL/6J) adult male testis cDNA to mRNA, clone.lib:RIKEN full-length enriched mouse cDNA library, clone:4933431F18.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 1842)  
AUTHORS Carninci,P. and Hayashizaki,Y.  
TITLE High-efficiency full-length cDNA cloning  
JOURNAL Methods in enzymology. 303, 19-44 (1999)  
MEDLINE 99279253  
PUBMED 10349536  
REFERENCE 2 (bases 1 to 1842)  
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes  
JOURNAL Genome research. 10 (10), 1617-1630 (2000)  
MEDLINE 20499374  
PUBMED 11042159  
REFERENCE 3 (bases 1 to 1842)  
AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsuunai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yanamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,





```
ORGANISM      Oryctolagus cuniculus
REFERENCE      Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS        1 (bases 1 to 544)
                Nikolaev,A.I., Ivanov,D.S., Amchenkova,A.M., Narovlyansky,A.N.,
                Panasyuk,A.F. and Turpaev,K.T.
TITLE          Isolation of nitric oxide-induced genes from chondrocyte cDNA
JOURNAL        Unpublished (2000)
COMMENT        Library using subtractive hybridization
                Contact: Turpaev K
                Engelhardt Institute of Molecular Biology
                Vavilova 32, Moscow, 117984, Russia
                Email: turpaev@genome.elmb.relarn.ru.
FEATURES       source
                1..544
                /organism="Oryctolagus cuniculus"
                /db_xref="taxon:9986"
                /clone="61"
                /clone_lib="Oryctolagus cuniculus femoral head 2 months
                old"
                /tissue_type="femoral head"
                /cell_type="primary articular chondrocyte"
                /dev_stage="2 months old"
                /note="breed: New Zealand White; obtained via subtractive
                hybridization of chondrocytes preincubated with 0.5 mM
                nitroglutathione vs. untreated cells"
BASE COUNT     173 a 102 c 103 g 166 t
ORIGIN
Query Match    8.1%; Score 26; DB 10; Length 544;
Best Local Similarity 100.0%; Pred. No. 0.0044;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 25 ATGATAACCCAGAGATAGAAGGTGG 50
      |||||||
Db 335 ATGATAACCCAGAGATAGAAGGTGG 310

RESULT 4
LOCUS      AA633769          406 bp      mRNA      EST      06-MAR-1998
DEFINITION AC27d11.s1 Stratiagene ovary (#937217) Homo sapiens cDNA clone
IMAGE:857685 3' similar to contains Alu repetitive element; contains
element PIR5 repetitive element ;, mRNA sequence.
ACCESSION   AA633769
VERSION     AA633769.1 GI:2556983
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 406)
AUTHORS     Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S.,
            Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin
            J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B.,
            White,Y., Wylie,T., Waterston,R. and Wilson,R.
TITLE       WashU-NCI human EST Project
JOURNAL     Unpublished (1997)
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LNL; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Possible reversed clone: polyt not found
            Insert Length: 659 Std Error: 0.00
            Seq primer: -40m13 fwd. ET from Amersham
            High quality sequence stop: 406.
            Location/Qualifiers
FEATURES
source
1..550
/organism="Homo sapiens"
/db_xref="GDB:4594051"
/db_xref="taxon:9606"
/clone="IMAGE:564987"
/clone_lib="Stratiagene lung carcinoma 937218"
/tissue_type="lung carcinoma"
/cell_line="NCI-H69"
/dev_stage="cell line NCI-H69"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: lung; Vector: pBluescript SK-; Site_1: EcoRI
; Site_2: XhoI; Cloned unidirectionally, Primer: Oligo
dt. Small cell carcinoma cell line NCI-H69. Average
```

```
1..406
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:857685"
/clone_lib="Stratiagene ovary (#937217)"
/sex="female"
/dev_stage="49 year old"
/lab_host="SOLR cells (kanamycin resistant)"
/note="Organ: ovary; Vector: Bluescript SK; Site_1: EcoRI;
Site_2: XhoI; Cloned unidirectionally. Primer: Oligo dt.
Total ovary tissue, normal, caucasian. Average insert
size: 0.8 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5'
GAATTCGGCAGAG 3' -3' adaptor sequence: 5'
CTCGAGTTTTTTTTTTTTT 3'
BASE COUNT     93 a 105 c 120 g 88 t
ORIGIN
Query Match    6.2%; Score 20; DB 10; Length 406;
Best Local Similarity 100.0%; Pred. No. 7.2;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 30 AACCCAGAGAGATAGAAGGTG 49
      |||||||
Db 328 AACCCAGAGAGATAGAAGGTG 347

RESULT 5
LOCUS      AA129435          550 bp      mRNA      EST      27-NOV-1996
DEFINITION zn85f02.s1 Stratiagene lung carcinoma 937218 Homo sapiens cDNA clone
IMAGE:564987 3', mRNA sequence.
ACCESSION   AA129435
VERSION     AA129435.1 GI:1689201
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 550)
AUTHORS     Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B.,
            Chisoe,S., Dietrich,N., DuBuque,T., Favello,A., Gish,W., Hawkins
            M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Mardis,E., Moore
            B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T.,
            Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J., Travaskis,E.,
            Underwood,K., Wohlmann,P., Waterston,R., Wilson,R. and Marra,M.
TITLE       Generation and analysis of 280,000 human expressed sequence tags
JOURNAL     Genome Res. 6 (9), 807-828 (1996)
KEYWORDS    97044478
MEDLINE
COMMENT     Contact: Wilson RK
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: est@watson.wustl.edu
            This clone is available royalty-free through LNL; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Seq primer: -40m13 fwd. from Amersham
            High quality sequence stop: 353.
            Location/Qualifiers
FEATURES
source
1..550
/organism="Homo sapiens"
/db_xref="GDB:4594051"
/db_xref="taxon:9606"
/clone="IMAGE:564987"
/clone_lib="Stratiagene lung carcinoma 937218"
/tissue_type="lung carcinoma"
/cell_line="NCI-H69"
/dev_stage="cell line NCI-H69"
/lab_host="SOLR (kanamycin resistant)"
/note="Organ: lung; Vector: pBluescript SK-; Site_1: EcoRI
; Site_2: XhoI; Cloned unidirectionally, Primer: Oligo
dt. Small cell carcinoma cell line NCI-H69. Average
```

insert size: 1.0 kb; Uni-ZAP XR Vector; ~5' adaptor  
sequence: 5' GAATTCGGCAGAG 3' ~3' adaptor sequence: 5'  
CTCAGCTTTTITTTTTTTT 3'

BASE COUNT 155 a 113 c 127 g 151 t 4 others

ORIGIN

Query Match 6.2%; Score 20; DB 10; Length 550;  
Best Local Similarity 100.0%; Pred. No. 7.2;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 289 AGGAAGACCCACATTACAGG 308  
|||||

Db 302 AGGAAGACCCACATTACAGG 321

RESULT 6

AZ681898 849 bp DNA GSS 14-DEC-2000  
LOCUS ENTJ277R Entamoeba histolytica Sheared DNA Entamoeba histolytica  
DEFINITION genomic, DNA sequence.

ACCESSION AZ681898  
VERSION AZ681898.1 GI:11818939  
KEYWORDS GSS.  
SOURCE Entamoeba histolytica.  
ORGANISM Entamoeba histolytica  
Eukaryota; Entamoebidae; Entamoeba.  
REFERENCE 1 (bases 1 to 849)  
AUTHORS Loftus, B., Van Aken, S. and Fraser, C.  
TITLE Determination of clone end sequences from Entamoeba histolytica  
JOURNAL HMI:IMSS sheared DNA library  
COMMENT Unpublished (2000)  
Contact: Brendan J Loftus  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: enta@tigr.org  
Clones are derived from the Entamoeba histolytica HMI:IMSS sheared  
DNA library

Seq primer: M13-Reverse  
Class: Shotgun  
High quality sequence start: 15  
High quality sequence stop: 821.  
Location/Qualifiers

FEATURES  
source  
1..849  
/organism="Entamoeba histolytica"  
/strain="HMI:IMSS"  
/db\_xref="taxon:5759"  
/clone\_lib="Entamoeba histolytica Sheared DNA"  
/notes="Vector: pHO51; Site 1: Bst I; Constructed at The  
Institute for Genomic Research (TIGR), Rockville, MD.  
Genomic DNA isolated from broth cultures of E. histolytica  
using a method described by Clark and Diamond (Clark,  
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a  
method for isolate identification. Exp. Parasitol.  
77:450.). The DNA was mechanically sheared to give a  
tight size distribution (~2 kb). The v + i method used for  
the library construction is described in detail in Smith,  
H.O. and Venter, J.C. (Making small insert libraries for  
whole genome shotgun sequencing projects. In Genome  
Sequencing: A Practical Approach, eds. M. Vaudin and B.  
Barell, Oxford University Press, 1999)."

BASE COUNT 330 a 104 c 110 g 305 t

ORIGIN

Query Match 6.2%; Score 20; DB 13; Length 849;  
Best Local Similarity 100.0%; Pred. No. 7.1;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 AAATAATTGTAATGTTATTG 22  
|||||

|||||

Db 469 AAATAATTGTAATGTTATTG 488

RESULT 7

AZ682025/c

LOCUS ENTLR91TF Entamoeba histolytica Sheared DNA Entamoeba histolytica  
DEFINITION genomic, DNA sequence.

ACCESSION AZ682025  
VERSION AZ682025.1 GI:11819171  
KEYWORDS GSS.  
SOURCE Entamoeba histolytica.  
ORGANISM Entamoeba histolytica  
Eukaryota; Entamoebidae; Entamoeba.  
REFERENCE 1 (bases 1 to 850)  
AUTHORS Loftus, B., Van Aken, S. and Fraser, C.  
TITLE Determination of clone end sequences from Entamoeba histolytica  
JOURNAL HMI:IMSS sheared DNA library  
COMMENT Unpublished (2000)  
Contact: Brendan J Loftus  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: enta@tigr.org  
Clones are derived from the Entamoeba histolytica HMI:IMSS sheared  
DNA library

Seq primer: M13-Forward  
Class: Shotgun  
High quality sequence start: 28  
High quality sequence stop: 693.  
Location/Qualifiers

FEATURES  
source  
1..850  
/organism="Entamoeba histolytica"  
/strain="HMI:IMSS"  
/db\_xref="taxon:5759"  
/clone\_lib="Entamoeba histolytica Sheared DNA"  
/notes="Vector: pHO51; Site 1: Bst I; Constructed at The  
Institute for Genomic Research (TIGR), Rockville, MD  
Genomic DNA isolated from broth cultures of E. histolytica  
using a method described by Clark and Diamond (Clark,  
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a  
method for isolate identification. Exp. Parasitol.  
77:450.). The DNA was mechanically sheared to give a  
tight size distribution (~2 kb). The v + i method used for  
the library construction is described in detail in Smith,  
H.O. and Venter, J.C. (Making small insert libraries for  
whole genome shotgun sequencing projects. In Genome  
Sequencing: A Practical Approach, eds. M. Vaudin and B.  
Barell, Oxford University Press, 1999)."

BASE COUNT 311 a 95 c 113 g 331 t

ORIGIN

Query Match 6.2%; Score 20; DB 13; Length 850;  
Best Local Similarity 100.0%; Pred. No. 7.1;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 AAATAATTGTAATGTTATTG 22  
|||||

Db 305 AAATAATTGTAATGTTATTG 286

RESULT 8

AZ689826/c

LOCUS ENTJ42TF Entamoeba histolytica Sheared DNA Entamoeba histolytica  
DEFINITION genomic, DNA sequence.

ACCESSION AZ689826  
VERSION AZ689826.1 GI:11826972  
KEYWORDS GSS.

Query Match 6.2%; Score 20; DB 13; Length 850;  
Best Local Similarity 100.0%; Pred. No. 7.1;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 AAATAATTGTAATGTTATTG 22  
|||||

SOURCE Entamoeba histolytica.  
 ORGANISM Entamoeba histolytica  
 REFERENCE 1 (bases 1 to 893)  
 AUTHORS Loftus, B., Van Aken, S. and Fraser, C.  
 TITLE Determination of clone end sequences from Entamoeba histolytica  
 JOURNAL HMI: IMSS sheared DNA library  
 COMMENT Unpublished (2000)  
 CONTACT: Brendan J Loftus  
 Department of Eukaryotic Genomics  
 The Institute for Genomic Research  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: enta@tigr.org  
 Clones are derived from the Entamoeba histolytica HMI:IMSS sheared  
 DNA library  
 Class: Shotgun  
 Seq primer: M13-Forward  
 High quality sequence start: 8  
 High quality sequence stop: 529.  
 Location/Qualifiers  
 1. .893  
 /organism="Entamoeba histolytica"  
 /strain="HMI:IMSS"  
 /db\_xref="taxon:5759"  
 /clone\_lib="Entamoeba histolytica Sheared DNA"  
 /note="Vector: pHOSt1; Site\_1: Bst I; Constructed at The  
 Institute for Genomic Research (TIGR), Rockville, MD.  
 Genomic DNA isolated from broth cultures of E. histolytica  
 using a method described by Clark and Diamond (Clark,  
 C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a  
 method for isolate identification. Exp. Parasitol.  
 77:450.). The DNA was mechanically sheared to give a  
 tight size distribution (~2 kb). The v + i method used for  
 the library construction is described in detail in Smith,  
 H.O. and Venter, J.C. (Making small insert libraries for  
 whole genome shotgun sequencing projects. In Genome  
 Sequencing: A Practical Approach, eds. M. Vaubin and B.  
 Barrel, Oxford University Press, 1999)."  
 296 a 101 c 134 g 362 t  
 BASE COUNT  
 ORIGIN  
 Query Match 6.2%; Score 20; DB 13; Length 893;  
 Best Local Similarity 100.0%; Pred. No. 7.1;  
 Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 3 AAATAATTGTTAATGTTATTG 22  
 |||||  
 Db 562 AAATAATTGTTAATGTTATTG 543  
 RESULT 9  
 BE711384/c  
 LOCUS 259 bp mRNA EST 12-SEP-2000  
 DEFINITION PM2-HT0686-250500-002-d09 HT0686 Homo sapiens cDNA, mRNA sequence.  
 ACCESSION BE711384  
 VERSION BE711384.1 GI:10099649  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 259)  
 Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,  
 Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,  
 Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bata, G.S., Simpson, D.H.,  
 Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare  
 , M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and  
 Simpson, A.J.  
 TITLE Shotgun sequencing of the human transcriptome with ORF expressed  
 sequence tags

JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
 MEDLINE 20202663  
 COMMENT Contact: Simpson A.J.G.  
 Laboratory of Cancer Genetics  
 Ludwig Institute for Cancer Research  
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
 Brazil  
 Tel: +55-11-2704922  
 Fax: +55-11-2707001  
 Email: asimpson@ludwig.org.br  
 This sequence was derived from the FAPESP/LICR Human Cancer Genome  
 project. This entry can be seen in the following URL  
 (http://www.ludwig.org.br/scripts/gethtml2.pl?cl-st2-PM2-HT0686-250  
 500-002-d09&t3=2000-05-25&t4=1)  
 Seq primer: puc 18 forward  
 High quality sequence start: 14  
 High quality sequence stop: 259.  
 Location/Qualifiers  
 1..259  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="HT0686"  
 /dev\_stage="Adult"  
 /note="Organ: head\_neck; Vector: puc18; Site\_1: SmaI;  
 Site\_2: SmaI; A mini-library was made by cloning products  
 derived from ORESTES PCR (U.S. Letters Patent application  
 No. 196,716 - Ludwig Institute for Cancer Research)  
 profiles into the puc 18 vector. Reverse transcription of  
 tissue mRNA and cDNA amplification were performed under  
 low stringency conditions."  
 75 a 33 c 55 g 96 t  
 BASE COUNT  
 ORIGIN  
 Query Match 5.9%; Score 19; DB 10; Length 259;  
 Best Local Similarity 100.0%; Pred. No. 25;  
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 298 CACATTACAGGACCCCTCC 316  
 |||||  
 Db 109 CACATTACAGGACCCCTCC 91  
 RESULT 10  
 AW559472/c  
 LOCUS 615 bp mRNA EST 07-SEP-2000  
 DEFINITION EST314520 DSIR Medicago truncatula cDNA clone pDSIR-1905, mRNA  
 sequence.  
 ACCESSION AW559472  
 VERSION AW559472.1 GI:7204898  
 KEYWORDS EST.  
 SOURCE Medicago.  
 ORGANISM barrel medic.  
 Medicago truncatula  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;  
 Medicago.  
 1 (bases 1 to 615)  
 Fedorova, M., Pierson, B.L., Samac, D.A., Vance, C.P., Gantt, G.S., Peng  
 , H., Ellis, L., Town, C.D., Bowman, C.L., Craven, M.B., Hansen, T.S.,  
 Holt, I.E. and Fraser, C.M.  
 ESTs from roots of Medicago truncatula after inoculation with  
 Phytophthora medicaginis  
 Unpublished (1999)  
 CONTACT: Carroll P. Vance  
 Department of Agronomy and Plant Genetics  
 University of Minnesota  
 411 Borlaug Hall, 1991 Upper Buford Circle, St. Paul, MN 55108 USA  
 Tel: 612 625 5715  
 Fax: 651-649-5058  
 Email: vance004@maroon.tc.umn.edu  
 Minnesota EST name: N250758e ; TIGR sequence name: MTBA187TK ; More  
 information, including clone ordering, is available at. .

```

'http://chrysic.tamu.edu/medicago'
Seq primer: Skn0d (CTA gAA CTA gTg GAT CC).
Location/Qualifiers
source
1. .615
/organism="Medicago truncatula"
/cultivar="genotype A17"
/db_xref="taxon:3880"
/clone_lib="pDSIR-1905"
/tissue_type="DSIR"
/dev_stage="roots infected with Phytophthora medicaginis"
/lab_host="E. coli strain XL0LR"
XhoI: cDNA was prepared from polyA+ enriched RNA from roots harvested at 10 days post inoculation with Phytophthora medicaginis. The cDNA was directionally ligated into the Uni-ZAP XR vector from Stratagene and packaged using Gigapack III gold packaging extracts. Plasmids containing cDNA inserts were excised from the recombinant lambda-ZAP phage using Ex-Assist helper phage and propagated in XL0LR cells. Note: EST may be of fungal origin."
BASE COUNT 187 a 122 c 126 g 180 t
ORIGIN

Query Match 5.9%; Score 19; DB 10; Length 615;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 232 CCAGACATGGCACTCTG 250
|||||
Db 279 CCAGACATGGCACTCTG 261

RESULT 11
AZ083136/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
AZ083136 666 bp DNA GSS 08-MAY-2000
RPCI-23-34P9.TV RPCI-23 Mus musculus genomic clone RPCI-23-34P9,
DNA sequence.
AZ083136
GSS.
house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret
,B., Lewis,M., McGann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
and Fraser,C.M.
Mouse BAC End Sequences from Library RPCI-23
Unpublished (1999)
Other GSSs: RPCI-23-34P9.TJ
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC
library availability, please contact Pieter de Jong
(pieter@dejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
or from Resea ch Genetics (info@resgen.com). BAC end page:
http://www.tigr.org/tldb/bacends/mouse/bac_end_intro.html
Plate: 34 row: P column: 9
Seq primer: T7
Class: BAC ends.
Location/Qualifiers
source
1. .666
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone_lib="RPCI-23-34P9"
/sex="Female"
/lab_host="DH10B"
/note="Organ: Kidney/Brain; Vector: pBACE3.6; Site_1:
EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
brain genomic DNA was isolated and partially digested
with a combination of EcoRI and EcoRI Methyase. Size
selected DNA was cloned into the pBACE3.6 vector at the
EcoRI sites. The ligation products were transformed into
DH10B electrocompetent cells (BRL Life Technologies)."
BASE COUNT 194 a 117 c 88 g 267 t
ORIGIN

Query Match 5.9%; Score 19; DB 13; Length 666;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 27 GATAACCAGAGATAGAA 45
|||||
Db 646 GATAACCAGAGATAGAA 628

RESULT 12
AA230263/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
AA230263 136 bp mRNA EST 19-AUG-1997
nc13a04.r1 NCI_CGAP_Prl Homo sapiens cDNA clone IMAGE:1007982, mRNA
sequence.
AA230263
EST.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs.rem@nih.gov
Tissue procurement: W. Marston Linehan, M.D., Rodrigo Chuaqui, M.D.
Michael Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: David B. Krizman, Ph.D.
CDNA Library Arrayed by: Genome Systems Inc., Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www.bio.llnl.gov/bbrp/image/image.html
Insert Length: 740 Std Error: 0.00
Seq primer: -28ml3 rev1 ET from Amersham
High quality sequence stop: 91.
Location/Qualifiers
source
1. .136
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NCI_CGAP_Prl"
/sex="Male"
/dev_stage="45 years old"
/lab_host="DH10B"
/note="Vector: pAMP10; Site_1: NotI; Site_2: EcoRI; 1st
strand cDNA was primed with oligo(dT)17 on 50 ng of
DNase-treated, total cellular RNA obtained from 5,000-10
,000 microdissected, histologically normal prostate
epithelial cells. Double-stranded cDNA was ligated to
EcoRI adaptors, 5 cycles of PCR applied to the cDNA with

```

an adaptor-specific primer, and the resulting PCR product subcloned into pAMP10 by the UDG-cloning method (Life Technologies). Average insert size is 600 bp. NOTE: Not directionally cloned. This library was constructed by David Krimman."

BASE COUNT 57 a 15 c 17 g 47 t

Query Match 5.6%; Score 18; DB 10; Length 136;  
Best Local Similarity 100.0%; Pred. No. 88;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAAATAATTGTAATGTT 18  
|||||

Db 80 GAAATAATTGTAATGTT 63

RESULT 13  
BF662254/c 155 bp mRNA EST 20-DEC-2000  
LOCUS  
DEFINITION maa80e02.y1 Soares mouse 3NbMS Mus musculus cDNA clone  
IMAGE:3823179 5', mRNA sequence.

ACCESSION BF662254  
VERSION BF662254.1 GI:11927388  
KEYWORDS  
SOURCE EST.

ORGANISM

Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.

1 (bases 1 to 155)  
NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

TITLE Tumor Gene Index

JOURNAL Unpublished (1997)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: [cgaps-r@mail.nih.gov](mailto:cgaps-r@mail.nih.gov)

This clone is available royalty-free through LLNL; contact the IMAGE Consortium ([infoimage.llnl.gov](http://infoimage.llnl.gov)) for further information.

MG1:1459531

Seq primer: -40RP from Gibco.

Location/Qualifiers

FEATURES

source

1..155

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="IMAGE:3823179"

/clone\_lib="Soares mouse 3NbMS"

/sex="male"

/tissue\_type="Spleen"

/dev\_stage="4 weeks"

/lab\_host="DH10B"

/note="vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5',

TGTTACCAATCTGAAGTGGAGCGCGCTGTTTTTTTTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I

and Eco RI sites of the modified pT7T3 vector. RNA provided by Dr. Bertrand Jordan. Library went through three rounds of normalization, and was constructed by

Bento Soares and M. Fatima Bonaldo."

BASE COUNT 47 a 24 c 29 g 55 t

ORIGIN

Query Match 5.6%; Score 18; DB 11; Length 155;  
Best Local Similarity 100.0%; Pred. No. 88;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 167 TTCAAAATGTGAAGAAGT 184

|||||

Db 76 TTCAAAATGTGAAGAAGT 59

RESULT 14  
A1784154/c

LOCUS

DEFINITION

IMAGE:2125733 3', mRNA sequence.

ACCESSION A1784154

VERSION A1784154.1 GI:5325963

KEYWORDS EST.

SOURCE human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 164)

NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

TITLE Tumor Gene Index

JOURNAL Unpublished (1997)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: [cgaps-r@mail.nih.gov](mailto:cgaps-r@mail.nih.gov)

This clone is available royalty-free through LLNL; contact the

IMAGE Consortium ([infoimage.llnl.gov](http://infoimage.llnl.gov)) for further information.

Insert Length: 418 Std Error: 0.00

Seq primer: -40UP from Gibco

High quality sequence stop: 159.

Location/Qualifiers

source

1..164

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:2125733"

/clone\_lib="Soares\_NSF\_F8\_9W\_OT\_PA\_P\_S1"

/lab\_host="DH10B"

/note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with

a modified polylinker; Site 1: Not I; Site 2: Eco RI;

Equal amounts of plasmid DNA from five normalized

libraries were mixed, and ss circles were made in vitro. In

Following HAP purification, this DNA was used as tracer in

a subtractive hybridization reaction. The driver was

PCR-amplified cDNAs from pools of 5,000 clones made from

the same 5 libraries. The pools consisted of the following

libraries and cloneIDs: Soares NbHSF pool 1:

309384-310919, 323208-325895 Soares Nb2HP pool 1:

145032-147335, 147720-148103, 148872-149255, 15002 -

150407, 151176-152327 Soares Nb2HP-9W pool 1:

758280-760583, 772104-774407 Soares NbHPA pool 1:

304776-306311, 320136-322823, 326280-326663 Soares NbHOT

pool 1: 723720-726407, 739080-740999 Subtraction by Bento

Soares and M. Fatima Bonaldo."

BASE COUNT 69 a 19 c 19 g 57 t

ORIGIN

Query Match 5.6%; Score 18; DB 10; Length 164;

Best Local Similarity 100.0%; Pred. No. 87;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAAATAATTGTAATGTT 18

|||||

Db 83 GAAATAATTGTAATGTT 66

RESULT 15

LOCUS

DEFINITION

AV298638

AV298638

AV298638

AV298638

AV298638

AV298638

AV298638

AV298638

AV298638

AV298638

AV298638

AV298638

AV298638

AV298638

A1784154 164 bp mRNA EST 13-DEC-1999  
t88d03.x1 Soares\_NSF\_F8\_9W\_OT\_PA\_P\_S1 Homo sapiens cDNA clone  
IMAGE:2125733 3', mRNA sequence.

ACCESSION A1784154

VERSION A1784154.1 GI:5325963

KEYWORDS EST.

SOURCE human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 164)

NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

TITLE Tumor Gene Index

JOURNAL Unpublished (1997)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: [cgaps-r@mail.nih.gov](mailto:cgaps-r@mail.nih.gov)

This clone is available royalty-free through LLNL; contact the

IMAGE Consortium ([infoimage.llnl.gov](http://infoimage.llnl.gov)) for further information.

Insert Length: 418 Std Error: 0.00

Seq primer: -40UP from Gibco

High quality sequence stop: 159.

Location/Qualifiers

source

1..164

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:2125733"

/clone\_lib="Soares\_NSF\_F8\_9W\_OT\_PA\_P\_S1"

/lab\_host="DH10B"

/note="Organ: pooled; Vector: pT7T3D-Pac (Pharmacia) with

a modified polylinker; Site 1: Not I; Site 2: Eco RI;

Equal amounts of plasmid DNA from five normalized

libraries were mixed, and ss circles were made in vitro. In

Following HAP purification, this DNA was used as tracer in

a subtractive hybridization reaction. The driver was

PCR-amplified cDNAs from pools of 5,000 clones made from

the same 5 libraries. The pools consisted of the following

libraries and cloneIDs: Soares NbHSF pool 1:

309384-310919, 323208-325895 Soares Nb2HP pool 1:

145032-147335, 147720-148103, 148872-149255, 15002 -

150407, 151176-152327 Soares Nb2HP-9W pool 1:

758280-760583, 772104-774407 Soares NbHPA pool 1:

304776-306311, 320136-322823, 326280-326663 Soares NbHOT

pool 1: 723720-726407, 739080-740999 Subtraction by Bento

Soares and M. Fatima Bonaldo."

BASE COUNT 69 a 19 c 19 g 57 t

ORIGIN

Query Match 5.6%; Score 18; DB 10; Length 164;  
Best Local Similarity 100.0%; Pred. No. 87;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GAAATAATTGTAATGTT 18

|||||

Db 83 GAAATAATTGTAATGTT 66

RESULT 15

LOCUS

DEFINITION

AV298638

AV298638

AV298638

AV298638

AV298638

AV298638

AV298638

AV298638

ORGANISM

Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 177)  
Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Carninci,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N., Hirozane,T., Hori,F., Ishii,Y., Ishikawa,T., Itoh,M., Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N., Kojima,Y., Koya,S., Kusakabe,M., Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y., Owa,C., Ozawa,Y., Saito,H., Sano,M., Sato,K., Shibata,K., Shibata,Y., Shigemoto,Y., Shiraki,T., Sogabe,Y., Sugahara,Y., Suzuki,H., Suzuki,H., Takahashi,F., Tateo,M., Tominaga,N., Tsunoda,Y., Watahiki,A., Watanabe,S., Yamamura,T., Yasunishi,A., Yokota,T., Yoshiki,A., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.  
RIKEN Mouse ESTs (Konno,H., et al. 1999)  
Unpublished (1999)  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@gsr.riken.go.jp,  
URL: http://genome.gsc.riken.go.jp/  
Sasaki,N., Izawa,M., Wataniki,M., Ozawa,K., Tanaka,T., Yoneda,Y., Matsuura,S., Carninci,P., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.

TITLE  
JOURNAL  
COMMENT

Transcriptional sequencing: A method for DNA sequencing using RNA polymerase. Proc. Natl. Acad. Sci. U.S.A. 95 (7), 3455-3460 (1998)  
Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J., Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki,Y. and Hayashizaki,Y.  
Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999)  
Carninci,P. and Hayashizaki,Y.  
High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 19-44 (1999)  
Please visit our web site (http://genome.rtc.riken.go.jp) for further details.

FEATURES

source

Location/Qualifiers

1..177  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="5730459p20"  
/clone\_lib="RIKEN full-length enriched, 8 days embryo"  
/sex="mixed"  
/dev\_stage="8 days embryo"  
/lab\_host="DH105"  
/note="Site\_1: Sali; Site\_2: BamHI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5'  
GAGAGAGAGAGATCCAGAGCTCTTTTTTTTTTTTTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of subtraction to Ret = 100.0 Second strand cDNA was prepared with the primer adapter of sequence [5'  
GAGAGAGATTCGAGTTAATTAATTCCTCCCCCCCC 3']. cDNA was cloned into the XhoI and BamHI sites. Vector: a modified pBluescript KS(+) after bulk excision from Lambda  
FLC I. Cloning sites, 5' end: Sali; 3' end: BamHI."

BASE COUNT  
ORIGIN

39 a 43 c 32 g 63 t

Query Match 5.6%; Score 18; DB 10; Length 177;  
Best Local Similarity 100.0%; Pred. No. 87;

	Matches	18:	Conservative	0:	Mismatches	0:	Indels	0:	Gaps	0:
QY	167	TTCAAAATTGTGAAGAAGT	184							
Db	107	TTCAAAATTGTGAAGAAGT	90							

Search completed: March 30, 2002, 06:44:20  
Job time: 5283 sec



**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2002, 05:19:47 ; Search time 89.47 Seconds  
(without alignments)  
810.025 Million cell updates/sec

Title: US-09-095-478A-4  
Perfect score: 320  
Sequence: 1 GAAATTAATGTAATGTTAT.....ATTACAGACCCCTCTTGT 320

Scoring table: OLIGO\_NUC  
Gap 60.0 , Gapext 60.0

Searched: 351203 seqs, 113238999 residues

Word size : 0

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued\_Patents\_NA: \*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq: \*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq: \*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq: \*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq: \*  
5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq: \*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	33	10.3	33	4	US-09-081-345-9
2	33	10.3	33	4	US-09-081-345-9
3	18	5.6	1398	2	US-08-896-320-2
4	16	5.0	618	4	US-09-328-111-830
5	16	5.0	634	4	US-09-328-111-825
6	16	5.0	965	2	US-09-141-135-1
7	16	5.0	2040	3	US-08-581-148C-21
8	16	5.0	2445	4	US-09-298-367B-9
9	16	5.0	2539	3	US-08-581-148C-20
10	16	5.0	2829	3	US-08-911-321-1
11	16	5.0	2844	3	US-08-815-809-2
12	16	5.0	2856	2	US-08-816-155B-4
13	16	5.0	2856	3	US-09-079-587-4
14	16	5.0	3060	1	US-08-286-305A-6
15	16	5.0	3060	2	US-08-441-104A-6
16	16	5.0	3060	2	US-08-440-816A-6
17	16	5.0	3060	4	US-09-417-381A-6
18	16	5.0	3194	2	US-08-359-705B-1
19	16	5.0	3194	2	US-08-286-846A-1
20	16	5.0	3194	2	US-08-457-880A-1
21	16	5.0	3194	3	US-08-444-622A-1
22	16	5.0	3194	3	US-08-942-562-1
23	16	5.0	3194	3	US-09-156-923-1
24	16	5.0	3339	3	US-08-911-321-5
25	16	5.0	3707	1	US-08-271-454-1
26	16	5.0	3707	5	PCT-US95-08180-1
27	16	5.0	4524	2	US-08-845-998-7

28 16 5.0 4524 3 US-09-206-537-7 Sequence 7, Appli  
29 16 5.0 4524 4 US-09-430-854-7 Sequence 7, Appli  
30 16 5.0 4833 4 US-09-066-047-1 Sequence 1, Appli  
31 16 5.0 6628 3 US-08-815-809-3 Sequence 3, Appli  
32 16 5.0 6649 2 US-08-816-155B-5 Sequence 5, Appli  
33 16 5.0 6649 3 US-09-079-587-5 Sequence 5, Appli  
34 16 5.0 8501 4 US-09-298-367B-6 Sequence 6, Appli  
35 15 4.7 34 4 US-09-081-345-7 Sequence 7, Appli  
36 15 4.7 34 4 US-09-081-345-7 Sequence 8, Appli  
37 15 4.7 486 2 US-08-506-864A-5 Sequence 5, Appli  
38 15 4.7 486 2 US-08-851-968-5 Sequence 5, Appli  
39 15 4.7 536 2 US-08-341-568-1 Sequence 1, Appli  
40 15 4.7 536 2 US-08-911-020-1 Sequence 1, Appli  
41 15 4.7 548 1 US-08-469-667-6 Sequence 6, Appli  
42 15 4.7 548 5 PCT-US95-07289-6 Sequence 6, Appli  
43 15 4.7 569 1 US-08-620-467A-6 Sequence 6, Appli  
44 15 4.7 569 1 US-08-348-572-6 Sequence 6, Appli  
45 15 4.7 569 3 US-09-041-090B-6 Sequence 6, Appli

## ALIGNMENTS

RESULT 1  
US-09-081-345-9  
; Sequence 9, Application US/09081345  
; Patent No. 6228641  
; GENERAL INFORMATION:  
; APPLICANT: Bahija Jallal  
; APPLICANT: Gregory D. Plovman  
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF  
; TITLE OF INVENTION: PTF04 RELATED DISORDERS  
; NUMBER OF SEQUENCES: 18  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Lyon & Lyon  
; STREET: 633 West Fifth Street  
; STREET: Suite 4700  
; CITY: Los Angeles  
; STATE: California  
; COUNTRY: U.S.A.  
; ZIP: 90071-2066  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
; MEDIUM TYPE: Storage  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: IBM P.C. DOS 5.0  
; SOFTWARE: Fastseq for Windows 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/081,345  
; FILING DATE: Herewith  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/047,222  
; FILING DATE: May 20, 1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Warburg, Richard J.  
; REGISTRATION NUMBER: 32,327  
; REFERENCE/DOCKET NUMBER: 234/253  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (213) 489-1600  
; TELEFAX: (213) 955-0440  
; TELEX: 67-3510  
; INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 33 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; US-09-081-345-9

Query Match 10.3%; Score 33; DB 4; Length 33;  
Best Local Similarity 100.0%; Pred. No. 2e-08;

Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 279 CCGTATGTGAGGAGCCACATTACAGGACC 311  
|||||  
Db 1 CCGTATGTGAGGAGCCACATTACAGGACC 33

## RESULT 2

US-09-081-345-10/c  
; Sequence 10, Application US/09081345  
; Patent No. 6228641  
; GENERAL INFORMATION:  
; APPLICANT: Bahija Jallal  
; APPLICANT: Gregory D. Plowman  
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF  
; TITLE OF INVENTION: PTP04 RELATED DISORDERS  
; NUMBER OF SEQUENCES: 18  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Lyon & Lyon  
; STREET: 633 West Fifth Street  
; STREET: Suite 4700  
; CITY: Los Angeles  
; STATE: California  
; COUNTRY: U.S.A.  
; ZIP: 90071-2066  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
; MEDIUM TYPE: storage  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: IBM P.C. DOS 5.0  
; SOFTWARE: FASTSEQ for Windows 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/081,345  
; FILING DATE: Herewith  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 60/047,222  
; FILING DATE: May 20, 1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Warburg, Richard J.  
; REGISTRATION NUMBER: 32,327  
; REFERENCE/DOCKET NUMBER: 234/253  
; TELEPHONE: (213) 489-1600  
; TELEFAX: (213) 955-0440  
; TELEX: 67-3510  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 33 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-09-081-345-10

Query Match 10.3%; Score 33; DB 4; Length 33;  
Best Local Similarity 100.0%; Pred. No. 2e-08;  
Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 279 CCGTATGTGAGGAGCCACATTACAGGACC 311  
|||||  
Db 33 CCGTATGTGAGGAGCCACATTACAGGACC 1

## RESULT 3

US-08-896-320-2  
; Sequence 2, Application US/08896320  
; Patent No. 5871971  
; GENERAL INFORMATION:  
; APPLICANT: Hillman, Jennifer L.  
; APPLICANT: Lal, Preeti  
; APPLICANT: Corley, Neil C.  
; APPLICANT: Shah, Purvi

; TITLE OF INVENTION: HUMAN DEVELOPMENTALLY REGULATED  
; TITLE OF INVENTION: GTP-BINDING PROTEIN  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Incyte Pharmaceuticals, Inc.  
; STREET: 3174 Porter Drive  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94304  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FASTSEQ for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/896,320  
; FILING DATE: Herewith  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Billings, Lucy J.  
; REGISTRATION NUMBER: 36,749  
; REFERENCE/DOCKET NUMBER: PF-0344 US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-855-0555  
; TELEFAX: 415-845-4166  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1398 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; LIBRARY: CARDNOT01  
; CLONE: 281964  
US-08-896-320-2

Query Match 5.6%; Score 18; DB 2; Length 1398;  
Best Local Similarity 100.0%; Pred. No. 2.2;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 167 TTCAATTGTGAAGAAGT 184  
|||||  
Db 1148 TTCAATTGTGAAGAAGT 1165

## RESULT 4

US-09-328-111-830  
; Sequence 830, Application US/09328111  
; Patent No. 6262333  
; GENERAL INFORMATION:  
; APPLICANT: Endege, Wilson O.  
; APPLICANT: Steinmann, Kathleen E.  
; APPLICANT: Astle, Jon H.  
; APPLICANT: Burgess, Christopher C.  
; APPLICANT: Bushnell, Steven E.  
; APPLICANT: Carroll III, Eddie  
; APPLICANT: Catino, Theodore J.  
; APPLICANT: Derti, Adnan  
; APPLICANT: Ford, Donna M.  
; APPLICANT: Lewis, Marcia E.  
; APPLICANT: Monahan, John E.  
; APPLICANT: Schlegel, Robert  
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION  
; TITLE OF INVENTION: PRODUCTS  
; FILE REFERENCE: CCD-257 (US)  
; CURRENT APPLICATION NUMBER: US/09/328,111  
; CURRENT FILING DATE: 1999-06-08

; EARLIER APPLICATION NUMBER: US 60/088,801  
; EARLIER FILING DATE: 1998-06-10  
; NUMBER OF SEQ ID NOS: 850  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 830  
; LENGTH: 618  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(618)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-328-111-830

Query Match 5.0%; Score 16; DB 4; Length 618;  
Best Local Similarity 100.0%; Pred. No. 24;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 196 AGTCACTCTGTAAAC 211  
DB 517 agtcactctgtaaaac 532

## RESULT 5

; Sequence 825, Application US/09328111  
; Patent No. 6262333  
; GENERAL INFORMATION:  
; APPLICANT: Endege, Wilson O.  
; APPLICANT: Steinmann, Kathleen E.  
; APPLICANT: Astle, Jon H.  
; APPLICANT: Burgess, Christopher C.  
; APPLICANT: Bushnell, Steven E.  
; APPLICANT: Carroll III, Eddie  
; APPLICANT: Catino, Theodore J.  
; APPLICANT: Derti, Adnan  
; APPLICANT: Ford, Donna M.  
; APPLICANT: Lewis, Marcia E.  
; APPLICANT: Monahan, John E.  
; APPLICANT: Schlegel, Robert  
; TITLE OF INVENTION: NOVEL HUMAN GENES AND GENE EXPRESSION  
; FILE REFERENCE: CCD-257 (US)  
; CURRENT APPLICATION NUMBER: US/09/328,111  
; CURRENT FILING DATE: 1998-06-08  
; EARLIER APPLICATION NUMBER: US 60/088,801  
; EARLIER FILING DATE: 1998-06-10  
; NUMBER OF SEQ ID NOS: 850  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 825  
; LENGTH: 634  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(634)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-328-111-825

Query Match 5.0%; Score 16; DB 4; Length 634;  
Best Local Similarity 100.0%; Pred. No. 24;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 186 CACAGGAAGAGTCAC 201  
DB 160 CACAGGAAGAGTCAC 145

## RESULT 6

; Sequence 1, Application US/09141135

; Patent No. 5981729  
; GENERAL INFORMATION:  
; APPLICANT: CHUN, Jong Yoon  
; APPLICANT: LEE, Yong Hun  
; TITLE OF INVENTION: Transcription Factor Gene Induced by Water Deficit and Abscisi  
; FILE REFERENCE: 1942/31  
; CURRENT APPLICATION NUMBER: US/09/141,135  
; CURRENT FILING DATE: 1998-08-27  
; NUMBER OF SEQ ID NOS: 2  
; SOFTWARE: WordPerfect 6.1/Windows  
; SEQ ID NO 1  
; LENGTH: 965  
; TYPE: DNA  
; ORGANISM: Arabidopsis thaliana  
US-09-141-135-1

Query Match 5.0%; Score 16; DB 2; Length 965;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 165 ATTTCAAATGTGAAG 180  
DB 377 ATTTCAAATGTGAAG 362

## RESULT 7

; Sequence 21, Application US/08581148C  
; Patent No. 6060644  
; GENERAL INFORMATION:  
; APPLICANT: Schnable, Patrick S.  
; APPLICANT: Robertson, Donald S.  
; APPLICANT: Hansen, Joel D.  
; APPLICANT: Nikolau, Basil J.  
; APPLICANT: Xu, Xiaojie  
; APPLICANT: Xia, Yiji  
; TITLE OF INVENTION: ISOLATION AND USE OF CUTICULAR LIPID  
; TITLE OF INVENTION: GENES  
; NUMBER OF SEQUENCES: 30  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Leydig, Voit & Mayer, Ltd.  
; STREET: Two Prudential Plaza, Suite 4900  
; CITY: Chicago  
; STATE: IL  
; COUNTRY: US  
; ZIP: 60601-6780  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/581,148C  
; FILING DATE: 29-DEC-1995  
; CLASSIFICATION: 800  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Larcher, Carol  
; REGISTRATION NUMBER: 35243  
; REFERENCE/DOCKET NUMBER: 71380  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (312) 616-5600  
; TELEFAX: (312) 616-5700  
; INFORMATION FOR SEQ ID NO: 21:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2040 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: CDS

LOCATION: 270..716  
NAME/KEY: CDS  
LOCATION: 1060..1875  
US-08-581-148C-21

Query Match 5.0%; Score 16; DB 3; Length 2040;  
Best Local Similarity 100.0%; Pred. No. 25;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 133 TTTCAGATAACTCAGT 148  
|||||  
Db 1054 TTTCAGATAACTCAGT 1069

## RESULT 8

US-09-298-367B-9  
Sequence 9, Application US/09298367B  
Patent No. 6180112  
GENERAL INFORMATION:  
APPLICANT: Highlander, Sarah K.  
TITLE OF INVENTION: PASTEURILLA HAEMOLYTICA VACCINE  
FILE REFERENCE: BCM-03728  
CURRENT APPLICATION NUMBER: US/09/298,367B  
CURRENT FILING DATE: 1999-04-22  
PRIOR APPLICATION NUMBER: 08/834,455  
PRIOR FILING DATE: 1997-04-15  
NUMBER OF SEQ ID NOS: 11  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 9  
LENGTH: 2445  
TYPE: DNA  
ORGANISM: Pasteurella haemolytica  
US-09-298-367B-9

Query Match 5.0%; Score 16; DB 4; Length 2445;  
Best Local Similarity 100.0%; Pred. No. 26;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 239 ATGGCACTCTCGCTC 254  
|||||  
Db 1772 atggcactctcgctc 1787

## RESULT 9

US-08-581-148C-20  
Sequence 20, Application US/08581148C  
Patent No. 6060644  
GENERAL INFORMATION:  
APPLICANT: Schnable, Patrick S.  
APPLICANT: Robertson, Donald S.  
APPLICANT: Hansen, Joel D.  
APPLICANT: Nikolau, Basil J.  
APPLICANT: Xu, Xiaojie  
APPLICANT: Xia, Yiji  
TITLE OF INVENTION: ISOLATION AND USE OF CUTICULAR LIPID  
NUMBER OF SEQUENCES: 30  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Leydig, Voit & Mayer, Ltd.  
STREET: Two Prudential Plaza, Suite 4900  
CITY: Chicago  
STATE: IL  
COUNTRY: US  
ZIP: 60601-6780  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/581,148C  
FILING DATE: 29-DEC-1995  
CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: Larcher, Carol  
REGISTRATION NUMBER: 35243  
REFERENCE/DOCKET NUMBER: 71380  
TELEPHONE: (312) 616-5600  
TELEFAX: (312) 616-5700  
INFORMATION FOR SEQ ID NO: 20:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2539 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-581-148C-20

Query Match 5.0%; Score 16; DB 3; Length 2539;  
Best Local Similarity 100.0%; Pred. No. 26;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 133 TTTCAGATAACTCAGT 148  
|||||  
Db 1354 TTTCAGATAACTCAGT 1369

## RESULT 10

US-08-911-321-1/c  
Sequence 1, Application US/08911321  
Patent No. 6010703  
GENERAL INFORMATION:  
APPLICANT: Roger K. Maes and Stephen J. Spatz  
TITLE OF INVENTION: Recombinant Poxvirus  
TITLE OF INVENTION: Vaccine Against  
TITLE OF INVENTION: Feline Rhinotracheitis  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Ian C. McLeod  
STREET: 2190 Commons Parkway  
CITY: Okemos  
STATE: Michigan  
COUNTRY: USA  
ZIP: 48864  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 5.25 inch, 360 Kb  
MEDIUM TYPE: storage  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: Wordperfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/911,321  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/096,183  
FILING DATE: July 26, 1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Ian C. McLeod  
REGISTRATION NUMBER: 20,931  
REFERENCE/DOCKET NUMBER: MSU 4.1-166  
TELEPHONE: (517) 347-4100  
TELEFAX: (517) 347-4103  
TELEX: No. 6010703e  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2829  
TYPE: Nucleic Acid  
STRANDEDNESS: Double

TOPOLOGY: Linear  
MOLECULE TYPE: DNA  
DESCRIPTION: DNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Feline herpesvirus-1  
STRAIN: 1  
INDIVIDUAL ISOLATE: C-27  
CELL TYPE: N/A  
FEATURE:  
NAME/KEY:  
LOCATION:  
IDENTIFICATION METHOD: Sequencing  
OTHER INFORMATION: DNA and deduced  
OTHER INFORMATION: polypeptide

US-08-911-321-1

Query Match 5.0%; Score 16; DB 3; Length 2829;  
Best Local Similarity 100.0%; Pred. No. 26;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 165 ATTCAAAATTGTGAAG 180  
|||||  
Db 1235 ATTCAAAATTGTGAAG 1220

## RESULT 11

US-08-815-809-2/c  
Sequence 2, Application US/08815809  
Patent No. 6004777

## GENERAL INFORMATION:

APPLICANT: TARTAGLIA, James  
APPLICANT: GOEBEL, Scott J.  
APPLICANT: COX, William I.  
APPLICANT: GETTIG, Russell R.  
APPLICANT: PINCUS, Steven E.  
APPLICANT: PAOLETTI, ENZO

APPLICANT: JACOBS, Bertram L.

TITLE OF INVENTION: VECTORS HAVING ENHANCED EXPRESSION, AND METHODS OF  
TITLE OF INVENTION: MAKING AND USES THEREOF

FILE REFERENCE: 454310-3010

CURRENT APPLICATION NUMBER: US/08/815,809

CURRENT FILING DATE: 1997-03-12

NUMBER OF SEQ ID NOS: 23

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 2

LENGTH: 2844

TYPE: DNA

ORGANISM: Vaccinia virus

US-08-815-809-2

## Query Match

Best Local Similarity 100.0%; Score 16; DB 3; Length 2844;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 165 ATTCAAAATTGTGAAG 180  
|||||  
Db 1247 ATTCAAAATTGTGAAG 1232

## RESULT 12

US-08-816-155B-4/c  
Sequence 4, Application US/08816155B  
Patent No. 5990091

## GENERAL INFORMATION:

APPLICANT: TARTAGLIA, JAMES  
APPLICANT: COX, WILLIAM I.  
APPLICANT: GETTIG, RUSSELL R.  
APPLICANT: MARTINEZ, HECTOR  
APPLICANT: PAOLETTI, ENZO

APPLICANT: PINCUS, STEVEN E.  
TITLE OF INVENTION: VECTORS HAVING ENHANCED EXPRESSION, AND  
TITLE OF INVENTION: METHODS OF MAKING AND USES THEREOF  
NUMBER OF SEQUENCES: 48  
CORRESPONDENCE ADDRESS:

ADDRESSEE: FROMMER LAWRENCE & HAUG LLP  
STREET: 745 FIFTH AVENUE  
CITY: NEW YORK  
STATE: NEW YORK  
COUNTRY: USA  
ZIP: 10151

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

## CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/816,155B

FILING DATE: 12-MAR-1997

CLASSIFICATION: 514

## ATTORNEY/AGENT INFORMATION:

NAME: KOWALSKI, THOMAS J.

REGISTRATION NUMBER: 32,147

REFERENCE/DOCKET NUMBER: 454310-2990

## TELECOMMUNICATION INFORMATION:

TELEPHONE: 212-588-0800

TELEFAX: 212-588-0500

INFORMATION FOR SEQ ID NO: 4:

## SEQUENCE CHARACTERISTICS:

LENGTH: 2856 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

US-08-816-155B-4

## Query Match

Best Local Similarity 100.0%; Score 16; DB 2; Length 2856;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 165 ATTCAAAATTGTGAAG 180  
|||||  
Db 1259 ATTCAAAATTGTGAAG 1244

## RESULT 13

US-09-079-587-4/c

Sequence 4, Application US/09079587

Patent No. 6130066

## GENERAL INFORMATION:

APPLICANT: TARTAGLIA, JAMES

APPLICANT: COX, WILLIAM I.

APPLICANT: GETTIG, RUSSELL R.

APPLICANT: MARTINEZ, HECTOR

APPLICANT: PAOLETTI, ENZO

APPLICANT: PINCUS, STEVEN E.

TITLE OF INVENTION: VECTORS HAVING ENHANCED EXPRESSION, AND  
TITLE OF INVENTION: METHODS OF MAKING AND USES THEREOF  
NUMBER OF SEQUENCES: 48

## CORRESPONDENCE ADDRESS:

ADDRESSEE: FROMMER LAWRENCE & HAUG LLP  
STREET: 745 FIFTH AVENUE  
CITY: NEW YORK  
STATE: NEW YORK  
COUNTRY: USA  
ZIP: 10151

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

## CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/079,587  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/816,155  
FILING DATE: 12-MAR-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: KOWALSKI, THOMAS J.  
REGISTRATION NUMBER: 32,147  
REFERENCE/DOCKET NUMBER: 454310-2990  
TELEPHONE: 212-588-0800  
TELEFAX: 212-588-0500  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2856 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-09-079-587-4

Query Match 5.0% Score 16; DB 3; Length 2856;  
Best Local Similarity 100.0% Pred No. 26;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 165 ATTCAAAATTGTGAAG 180  
|||||  
DB 1259 ATTCAAAATTGTGAAG 1244

## RESULT 14

US-08-286-305A-6  
Sequence 6, Application US/08286305A  
Patent No. 5766863

GENERAL INFORMATION:  
APPLICANT: Godowski, Paul J.  
APPLICANT: Mark, Melanie R.  
APPLICANT: Sadick, Michael D.  
APPLICANT: Shelton, David L.  
APPLICANT: Wong, Wai Lee Tan  
TITLE OF INVENTION: KINASE RECEPTOR ACTIVATION ASSAY  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 460 Point San Bruno Blvd  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080

COMPUTER READABLE FORM:  
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: patin (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/286,305A  
FILING DATE: 05-AUG-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/170558  
FILING DATE: 20-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/157563  
FILING DATE: 23-NOV-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Lee, Wendy M.  
REGISTRATION NUMBER: 00,000  
REFERENCE/DOCKET NUMBER: 854C1PI  
TELEPHONE: 415/225-1994  
TELEFAX: 415/952-9881

TELEX: 910/371-7168  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3060 bases  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-286-305A-6

Query Match 5.0% Score 16; DB 1; Length 3060;  
Best Local Similarity 100.0% Pred No. 26;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 87 GAAGGAGCCTTTGGAA 102  
|||||  
DB 2017 GAAGGAGCCTTTGGAA 2032

## RESULT 15

US-08-441-104A-6  
Sequence 6, Application US/08441104A  
Patent No. 5891650

GENERAL INFORMATION:  
APPLICANT: Godowski, Paul J.  
APPLICANT: Mark, Melanie R.  
APPLICANT: Sadick, Michael D.  
APPLICANT: Shelton, David L.  
APPLICANT: Wong, Wai Lee Tan  
TITLE OF INVENTION: KINASE RECEPTOR ACTIVATION ASSAY  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 460 Point San Bruno Blvd  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080

COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch, 720 kb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: patin (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/441,104A  
FILING DATE: 15-MAY-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/286305  
FILING DATE: 05-AUG-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/170558  
FILING DATE: 20-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/157563  
FILING DATE: 23-NOV-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Lee, Wendy M.  
REGISTRATION NUMBER: 00,000  
REFERENCE/DOCKET NUMBER: 854C1PIC2  
TELEPHONE: 415/225-1994  
TELEFAX: 415/952-9881  
TELEX: 910/371-7168  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3060 bases  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
US-08-441-104A-6

Query Match 5.0%; Score 16; DB 2; Length 3060;  
Best Local Similarity 100.0%; Pred. No. 26;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 87 GAAGGAGCCTTTGGAA 102  
|||||  
Db 2017 GAAGGAGCCTTTGGAA 2032

Search completed: March 30, 2002, 07:10:46  
Job time: 6659 sec

**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2002, 05:16:48 ; Search time 1463.6 Seconds  
(without alignments)  
3606.926 Million cell updates/sec

Title: US-09-095-478A-4  
Perfect score: 320  
Sequence: 1 GAATAATTTGTAATGTTAT.....ATTACAGGACCCCTCTTGT 320

Scoring table:  
OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 1472140 seqs, 8248589755 residues

Word size : 0

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb.ba.\*
- 2: gb.htg.\*
- 3: gb.in.\*
- 4: gb.om.\*
- 5: gb.ov.\*
- 6: gb.pat.\*
- 7: gb.ph.\*
- 8: gb.pl.\*
- 9: gb.pr.\*
- 10: gb.ro.\*
- 11: gb.sts.\*
- 12: gb.sy.\*
- 13: gb.un.\*
- 14: gb.vi.\*
- 15: em.ba.\*
- 16: em.fun.\*
- 17: em.hum.\*
- 18: em.in.\*
- 19: em.om.\*
- 20: em.or.\*
- 21: em.ov.\*
- 22: em.pat.\*
- 23: em.ph.\*
- 24: em.pl.\*
- 25: em.ro.\*
- 26: em.sts.\*
- 27: em.sy.\*
- 28: em.un.\*
- 29: em.vi.\*
- 30: em.htgo.hum.\*
- 31: em.htgo.inv.\*
- 32: em.htgo.rod.\*
- 33: em.htg.hum.\*
- 34: em.htg.inv.\*
- 35: em.htg.rod.\*
- 36: em.htg.other.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

8

Result No.	Score	Match	Length	DB	ID	Description
1	50	15.6	3090	10	D64141	D64141 Mus musculus
2	33	10.3	33	6	AR149921	AR149921 Sequence
3	33	10.3	33	6	AR149922	AR149922 Sequence
4	28	8.8	2786	9	HSN800374	AL500040 Homo sapi
5	28	8.8	197389	9	AL558791	AL558791 Human DNA
6	21	6.6	137584	2	AC005845	AC005845 Homo sapi
7	21	6.2	95241	9	HS39819	AL023096 Human DNA
8	20	6.2	148269	2	AC026486	AC026486 Homo sapi
9	20	6.2	158312	2	AC083922	AC083922 Homo sapi
10	20	6.2	162249	2	AC020682	AC020682 Homo sapi
11	20	6.2	166973	9	AL451049	AL451049 Human DNA
12	20	6.2	188026	2	AL356952	AL356952 Homo sapi
13	20	6.2	273403	2	AC011498	AC011498 Homo sapi
14	19	5.9	1254	8	AB011796	AB011796 Citrus un
15	19	5.9	8935	14	AF179368	AF179368 HIV-1 str
16	19	5.9	79262	8	AC006587	AC006587 Arabidops
17	19	5.9	86629	9	AF305873	AF305873 Homo sapi
18	19	5.9	86867	9	AF230666	AF230666 Homo sapi
19	19	5.9	89122	9	AC007199	AC007199 Homo sapi
20	19	5.9	97168	9	AF230667	AF230667 Homo sapi
21	19	5.9	97693	9	AC002383	AC002383 Homo sapi
22	19	5.9	125066	9	AC074286	AC074286 Homo sapi
23	19	5.9	131611	9	AC005099	AC005099 Homo sapi
24	19	5.9	148245	2	AC016209	AC016209 Homo sapi
25	19	5.9	153000	2	AC010934	AC010934 Homo sapi
26	19	5.9	162098	2	AC019032	AC019032 Homo sapi
27	19	5.9	167880	9	HU091323	U91323 Human Chrom
28	19	5.9	171374	9	AC010980	AC010980 Homo sapi
29	19	5.9	171769	8	FLN3	AC007519 Sequence
30	19	5.9	179441	2	AC044879	AC044879 Homo sapi
31	19	5.9	179728	2	AC087820	AC087820 Homo sapi
32	19	5.9	180347	9	AC009415	AC009415 Homo sapi
33	19	5.9	183610	2	AC026452	AC026452 Homo sapi
34	19	5.9	198740	2	AC069434	AC069434 Homo sapi
35	19	5.9	211607	2	AC025778	AC025778 Homo sapi
36	18	5.6	332	6	AX150131	AX150131 Sequence
37	18	5.6	718	9	HSN801361	AL133085 Homo sapi
38	18	5.6	1380	10	AF232010	AF232010 Cavla sp.
39	18	5.6	1383	9	HSN805940	AJ005940 Homo sapi
40	18	5.6	1398	6	AR035970	AR035970 Sequence
41	18	5.6	1477	9	AF131829	AF131829 Homo sapi
42	18	5.6	1647	9	AF078103	AF078103 Homo sapi
43	18	5.6	1666	11	G06702	G06702 human STS W
44	18	5.6	1910	8	CBUS3145	U53145 Candida boi
45	18	5.6	2137	8	YSAPMP47	J05672 Candida boi

ALIGNMENTS

RESULT 1

D64141	Mus musculus	3090 bp	mRNA	ROD	13-FEB-1999
LOCUS	Mus musculus mRNA for protein-tyrosine-phosphatase, complete cds.				
DEFINITION	D64141				
ACCESSION	D64141.1 GI:2665457				
VERSION	protein-tyrosine-phosphatase.				
KEYWORDS	Mus musculus testis cDNA to mRNA.				
SOURCE	Mus musculus				
ORGANISM	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
REFERENCE	1 (bases 1 to 3090)				
AUTHORS	Ohsugi, M.				
TITLE	Direct Submission				
JOURNAL	Submitted (15-SEP-1995) to the DDBJ/EMBL/GenBank databases. Miho Ohsugi, The University of Tokyo, Department of Oncology; 4-6-1 Shirokanedai, Minato-ku, Tokyo 108, Japan (E-mail:mohsugi@ims.u-tokyo.ac.jp, Tel:03-5449-5305, Fax:03-5449-5413)				
REFERENCE	2 (sites)				

AUTHORS Obsugi, M., Kuramochi, S., Matsuda, S. and Yamamoto, T.  
 TITLE Molecular cloning and characterization of a novel cytoplasmic protein-tyrosine phosphatase that is specifically expressed in spermatocytes  
 JOURNAL J. Biol. Chem. 272 (52), 33092-33099 (1997)  
 MEDLINE 98070510  
 FEATURES Location/Qualifiers  
 source 1..3090  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /tissue\_type="testis"  
 218..1498  
 /EC\_number="3.1.3.48"  
 /codon\_start=1  
 /product="protein-tyrosine-phosphatase"  
 /protein\_id="BAA23761.1"  
 /db\_xref="GI:2665458"  
 /translations="MSPRKVRKGTGRDNDDEECNSGNLRLNSLPSSSKOKMPTTPI  
 FGKMSNENKPSHLSFSDKYELVPEPLESDTDETVMDVDSRLNRNWSMDSETA  
 GPSKTVSPVLUGSSRLSKDPTSVSEKELTQIAQIRPLIFNSARSAMRDLNTLQK  
 EELDIREFLEQMTLPDDFNSGNTLQNRDKNRDILFYDSTRVPLGKNKDYINAS  
 YIRIVHEEYFYIATQGPLPETIEDFQWVLENNCNVIAITRETCGVKCYSWP  
 ISLKEPLERHFSVLETRHTVTOYFTVRVQIVKSTGSKOCVKHLQFTKWPDPGPA  
 SADFFIKYRYVRKSHITGPLLHVSACVGRVTCFVCDVWFSAIEKNYSFDLMNIVT  
 OMRRQKGMITQREQFCYEVLEVLQNLALY"  
 BASE COUNT 904 a 603 c 647 g 936 t  
 ORIGIN

Query Match 15.6%; Score 50; DB 10; Length 3090;  
 Best Local Similarity 100.0%; Pred. No. 2.9e-16;  
 Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 271 AAATATGTCCTTATGTGAGGAAGAGCCACATTACAGGACCCCTCTGT 320  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 1238 AAATATGTCCTTATGTGAGGAAGAGCCACATTACAGGACCCCTCTGT 1287

RESULT 2  
 ARI49921  
 LOCUS ARI49921 33 bp DNA PAT 08-AUG-2001  
 DEFINITION Sequence 9 from patent US 6228641.  
 ACCESSION ARI49921  
 VERSION ARI49921.1 GI:15114512  
 KEYWORDS Unknown.  
 SOURCE Unknown.  
 ORGANISM Unknown.  
 REFERENCE 1 (bases 1 to 33)  
 AUTHORS Jallal, B. and Plozman, G.D.  
 TITLE Diagnosis and treatment of PTP04 related disorders  
 JOURNAL Patent: US 6228641-A 9 08-MAY-2001;  
 FEATURES Location/Qualifiers  
 source 1..33  
 /organism="unknown"  
 BASE COUNT 10 a 8 c 9 g 6 t  
 ORIGIN

Query Match 10.3%; Score 33; DB 6; Length 33;  
 Best Local Similarity 100.0%; Pred. No. 5.6e-07;  
 Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 279 CCGTTATGTGAGGAAGAGCCACATTACAGGACC 311  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 1 CCGTTATGTGAGGAAGAGCCACATTACAGGACC 33

RESULT 3  
 ARI49922/c  
 LOCUS ARI49922 33 bp DNA PAT 08-AUG-2001  
 DEFINITION Sequence 10 from patent US 6228641.  
 ACCESSION ARI49922

ARI49922.1 GI:15114513  
 KEYWORDS Unknown.  
 SOURCE Unknown.  
 ORGANISM Unknown.  
 REFERENCE 1 (bases 1 to 33)  
 AUTHORS Jallal, B. and Plozman, G.D.  
 TITLE Diagnosis and treatment of PTP04 related disorders  
 JOURNAL Patent: US 6228641-A 10 08-MAY-2001;  
 FEATURES Location/Qualifiers  
 source 1..33  
 /organism="unknown"  
 BASE COUNT 6 a 9 c 8 g 10 t  
 ORIGIN

Query Match 10.3%; Score 33; DB 6; Length 33;  
 Best Local Similarity 100.0%; Pred. No. 5.6e-07;  
 Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 279 CCGTTATGTGAGGAAGAGCCACATTACAGGACC 311  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 33 CCGTTATGTGAGGAAGAGCCACATTACAGGACC 1

RESULT 4  
 HSM800374  
 LOCUS HSM800374 2786 bp mRNA PRI 18-FEB-2000  
 DEFINITION Homo sapiens mRNA; cDNA DKFp566K0524 (from clone DKFp566K0524); partial cds.  
 ACCESSION AL050040  
 VERSION AL050040.1 GI:4884281  
 KEYWORDS human.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 REFERENCE 1 (bases 1 to 2786)  
 AUTHORS Ansoorge, W., Wirkner, U., Mewes, H.W., Gassenhuber, J. and Wiemann, S.  
 TITLE Direct Submission  
 JOURNAL Submitted (15-MAY-1999) MIPS, Am Klopferspitze 18a, D-82152 Martinsried, GERMANY  
 COMMENT Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de; sequenced by EMBL (European Molecular Biology Laboratories, Heidelberg/Germany) within the cDNA sequencing consortium of the German Genome Project.  
 This clone (DKFp566K0524) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further information about the clone and the sequencing project is available at http://www.mips.biochem.mpg.de/proj/cDNA/.

FEATURES Location/Qualifiers  
 source 1..2786  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="DKFp566K0524"  
 /clone\_lib="566 (synonym: hfk42). Vector pAMP1; host XL-2blue; sites NotI + SalI"  
 /dev\_stage="fetal"  
 /tissue\_type="kidney"  
 1..1198  
 /gene="DKFp566K0524"  
 <1..1198  
 /gene="DKFp566K0524"  
 /note="strong similarity to protein-tyrosine-phosphatases"  
 /codon\_start=2  
 /product="hypothetical protein"  
 /protein\_id="CAB43248.1"  
 /db\_xref="GI:4884282"  
 /translation="GNWRGCGSCCTVGAAPQAVFENKVNSEKVKLSIRNPHNDY  
 EDVFEPSGSGSDPSMWTARGPRRDRNSEDDEAAGPSQALSFLLSDTRKIVSEGEL



```

repeat_region 21217..21621
/note="MLT1A1-internal repeat: matches 927..1333 of
consensus"
repeat_region 22088..22592
/note="MLT1A1-internal repeat: matches 160..712 of
consensus"
repeat_region 22617..22732
/note="L1ME3 repeat: matches 6022..6145 of consensus"
repeat_region 22795..22911
/note="MLT1A1-internal repeat: matches 5..121 of
consensus"
repeat_region 22917..23273
/note="MLT1A1 repeat: matches 1..365 of consensus"
repeat_region 23274..23344
/note="MLT1I repeat: matches 339..410 of consensus"
repeat_region 24336..24385
/note="MLN1F repeat: matches 268..317 of consensus"
repeat_region 25059..26364
/note="L1M4 repeat: matches 1064..2552 of consensus"
repeat_region 26398..26843
/note="L1M4 repeat: matches 2651..3102 of consensus"
repeat_region 26879..27248
/note="L1ME3 repeat: matches 5551..5956 of consensus"
repeat_region 27255..27500
/note="MLT1F repeat: matches 304..568 of consensus"
repeat_region 27667..27736
/note="MLT1F repeat: matches 3..74 of consensus"
repeat_region 27759..27927
/note="L1ME3 repeat: matches 5992..6162 of consensus"
repeat_region 30024..30310
/note="L1ME3 repeat: matches 1151..1435 of consensus"
repeat_region 31602..31678
/note="MIR repeat: matches 79..155 of consensus"
misc_feature 31627..31927
/note="match: STS: Em:HSJ10C10"
repeat_region 31786..31900
/note="MIR repeat: matches 89..203 of consensus"
repeat_region 32440..32738
/note="AluX repeat: matches 1..300 of consensus"
repeat_region 32827..32870
/note="22 copies 2 mer tt 75% conserved"
repeat_region 34892..35012
/note="L2 repeat: matches 2584..2710 of consensus"
repeat_region 36225..36549
/note="MSRD repeat: matches 21..394 of consensus"
repeat_region 36732..36915
/note="MIR repeat: matches 36..234 of consensus"
repeat_region 37945..38060
/note="MIR repeat: matches 48..170 of consensus"
repeat_region 38422..38524
/note="MER5A repeat: matches 9..116 of consensus"
repeat_region 38641..38785
/note="MIR repeat: matches 34..192 of consensus"
repeat_region 39515..39846
/note="L1ME3 repeat: matches 5845..6182 of consensus"
repeat_region 40361..41030
/note="MLT1H repeat: matches 72..547 of consensus"
repeat_region 41704..41789
/note="MIR repeat: matches 84..166 of consensus"
repeat_region 41923..42236
/note="L1ME1 repeat: matches 5808..6163 of consensus"
repeat_region 45114..45484
/note="MLT1A1 repeat: matches 1..365 of consensus"
repeat_region 46495..46591
/note="MLT1J repeat: matches 43..140 of consensus"
repeat_region 47474..47516
/note="L2 repeat: matches 2658..2700 of consensus"
repeat_region 48316..48775
/note="L1M4 repeat: matches 6443..6923 of consensus"
repeat_region 49375..49685
/note="AluX repeat: matches 1..299 of consensus"
repeat_region 49686..49720
/note="L1R12 repeat: matches 445..479 of consensus"

```

```

repeat_region 49691..49724
/note="PTR5 repeat: matches 173..206 of consensus"
repeat_region 49725..50193
/note="L1R12 repeat: matches 19..1523 of consensus"

Query Match 8.8%; Score 28; DB 9; Length 197389;
Best Local Similarity 100.0%; Pred. No. 0.0003;
Matches 28; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Q# 229 TGGCCAGACCATGCGACTCTGCTCTCAG 256
Db 158585 TGGCCAGACCATGCGACTCTGCTCTCAG 158558
|||||
AC005845 137584 bp DNA HTG 17-JUL-2001
Homo sapiens chromosome 12 clone RP3-454B23, WORKING DRAFT
SEQUENCE, 5 unordered pieces.
AC005845
AC005845.8 GI:14787165
HTG: HTGS_GISEL; HTGS_DRAFT; HTGS_FULLTOP; HTGS_ACTIVEFIN.
human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-Osman, F.R., Allen, C.,
Alsbrooks, S.L., Amaratunga, H.C., Are, J.R., Banks, T., Barbara, J.,
Benton, J., Bimberg, K., Blankenburg, K., Bonnin, D., Bouck, J.,
Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C.,
Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C., Carron, T.F.,
Chen, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R.,
Chen, Z., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C.,
Coyle, M.D., Dathorne, S.R., David, R., Davila, M.L., Davis, C.,
Davy, Carroll, L., Dederich, D.A., Delaney, K.R., Delgado, O.,
Denn, A.L., Ding, Y., Dinh, H.H., Douthwaite, K.J., Draper, H.,
Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C.,
Elhaj, C., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J.,
Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T.,
Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S.,
Hernandez, J., Hernandez, O., Hodgson, A., Hogue, M., Holloway, C.,
Hollins, B., Homs, F., Howard, S., Huber, J., Hulyk, S., Hume, J.,
Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S.,
Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J.,
Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C.,
Lewis, L., Li, J., Li, Z., Lichtarge, O., Lieu, C., Liu, J., Liu, W.,
Loulsegh, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, K., Luna, R.,
Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A.,
Martinez, E., Massey, E., Mawhiney, E., McLeod, M.P., Meador, M.,
Mei, G., Metzger, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K.,
Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N.,
Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenwo, S.,
Ogih, M., Okunolu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B.,
Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L.,
Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojibokan, I., Rolfe, M.,
Ruiz, S., Saverly, G., Scherer, S., Scott, G., Shen, H., Shoohtari, N.,
Sisson, I., Sodergren, E., Sonaike, T., Sparks, A., Stanley, H.,
Stone, H., Sutton, A., Svatek, A., Tabor, P., Tameria, A., Tameria, K.,
Tang, H., Tansey, J., Taylor, C., Taylor, T., Teifrod, B., Thomas, N.,
Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R.,
Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C.,
Watlington, S., Williams, G., Williamson, A., Wleczky, R., Wooden, S.,
Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
Weinstock, G. and Gibbs, R.

Direct Submission
Unpublished
2 (bases 1 to 137584)
Direct Submission
Submitted (22-OCT-1998) Molecular and Human Genetics, Baylor
College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

```

## COMMENT

On Jul 17, 2001 this sequence version replaced gi:14625075.  
----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: ZR  
Center clone name: RP3-454B23  
----- Summary Statistics  
Sequencing vector: Plasmid; M7789  
Chemistry: Dye-terminator Big Dye; 4% of reads  
Assembly program: Phrap; version 0.990329  
Consensus quality: 139066 bases at least Q40  
Consensus quality: 142794 bases at least Q30  
Consensus quality: 144767 bases at least Q20  
Estimated insert size: 137687; sum-of-contigs estimation  
Quality coverage: 10.8x in Q20 bases; sum-of-contigs estimation  
-----

\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 5 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 58304: contig of 58304 bp in length  
\* 58305 58404: gap of unknown length  
\* 58405 97131: contig of 38727 bp in length  
\* 97132 97231: gap of unknown length  
\* 97232 111131: contig of 13900 bp in length  
\* 111132 111231: gap of unknown length  
\* 111232 126332: contig of 15101 bp in length  
\* 126333 126432: gap of unknown length  
\* 126433 137584: contig of 11152 bp in length.

## FEATURES

source

1..137584  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="12"  
/clone="RP3-454B23"  
BASE COUNT 36835 a 33464 c 33266 g 33618 t 401 others  
ORIGIN

Query Match 6.6%; Score 21; DB 2; Length 137584;  
Best Local Similarity 100.0%; Pred. No. 2;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 103 TTCAACACTTTCATGTCCTT 123  
|||||  
Db 5514 TTCAACACTTTCATGTCCTT 5494

## RESULT

HS39819

LOCUS HS39819 95241 bp DNA PRI 23-NOV-1999  
DEFINITION Human DNA sequence from clone 39819 on chromosome 1p34.3-36.13  
Contains acute myeloid leukaemia (AML1) gene, EST, GSS, CpG island,  
complete sequence.

ACCESSION

AL023096

VERSION

AL023096.1

KEYWORDS

HTG; acute myeloid leukaemia; AML1; CpG island; runt domain;

transcription factor.

SOURCE

Homo sapiens

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

## REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

1 (bases 1 to 95241)

Grafham,D.

Direct Submission

Submitted (04-AUG-1998) E-mail enquiries: humquery@sanger.ac.uk

Clone requests: clonerequest@sanger.ac.uk

On Nov 21, 1998 this sequence version replaced gi:3114715.

During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.

This sequence is the entire insert of clone 39819. This sequence  
has been finished according to sequence map criteria as follows. An  
attempt is made to resolve all sequencing problems, such as  
compressions and repeats, but not necessarily within known  
annotated human repeat sequence elements (e.g. Alu). Where the  
sequence is ambiguous, there is an annotation using the 'unsure'  
feature key.

This sequence was generated from part of bacterial clone contigs of  
human chromosome 1, constructed by the Sanger Centre Chromosome 1  
Mapping Group. Further information can be found at  
<http://www.sanger.ac.uk/HGP/Chr1>

39819 is from the library RPCI3 constructed at the Roswell Park

Cancer Institute by the group of Pieter de Jong. For further

details see <http://bacpac.med.buffalo.edu/VECTOR:pcypac2>.

## FEATURES

source

1..95241

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="1"

/map="p34.3-36.13"

/clone="RP3-39819"

/clone\_lib="RPCI-3"

1..242

repeat\_region

/note="AluX repeat: matches 59..302 of consensus"

555..634

repeat\_region

/note="MIR repeat: matches 69..147 of consensus"

849..1014

repeat\_region

/note="AluSx repeat: matches 1..168 of consensus"

1022..1249

repeat\_region

/note="AluJo repeat: matches 85..301 of consensus"

complement(1624..1743)

repeat\_region

/note="MIR repeat: matches 151..33 of consensus"

1888..2058

repeat\_region

/note="MIR repeat: matches 21..195 of consensus"

complement(2098..2239)

repeat\_region

/note="MIR2 repeat: matches 146..4 of consensus"

3109..3175

repeat\_region

/note="MIR repeat: matches 73..139 of consensus"

3924..4206

repeat\_region

/note="AluSg repeat: matches 1..298 of consensus"

4270..4481

repeat\_region

/note="MER42c repeat: matches 1248..1481 of consensus"

complement(4490..4789)

repeat\_region

/note="AluSx repeat: matches 301..2 of consensus"

complement(4790..4915)

repeat\_region

/note="AluSx repeat: matches 136..1 of consensus"

complement(5362..5662)

repeat\_region

/note="AluSg repeat: matches 303..1 of consensus"

5918..6029

repeat\_region

/note="MIR repeat: matches 35..147 of consensus"

complement(7365..7572)

repeat\_region

/note="MIR repeat: matches 240..14 of consensus"

complement(8168..8279)

repeat\_region

/note="MIR repeat: matches 192..80 of consensus"

8835..9131

repeat\_region

/note="AluSx repeat: matches 3..302 of consensus"

9426..9721

repeat\_region

/note="AluJo repeat: matches 1..300 of consensus"

complement(10746..11046)

repeat\_region

/note="AluSx repeat: matches 302..1 of consensus"

11707..11857

repeat\_region

```

repeat_region /note="MIR repeat: matches 34. .179 of consensus"
11928. .12036
/note="MER5A repeat: matches 2. .108 of consensus"
repeat_region complement(12047. .12348)
/note="AluX repeat: matches 302. .1 of consensus"
repeat_region complement(12368. .12685)
/note="L1MB7 repeat: matches 907. .574 of consensus"
repeat_region complement(12783. .13172)
/note="L1MC1 repeat: matches 1063. .672 of consensus"
repeat_region complement(13175. .13475)
/note="MSTD repeat: matches 392. .82 of consensus"
repeat_region complement(13478. .13776)
/note="AluJo repeat: matches 300. .1 of consensus"
repeat_region complement(13777. .13869)
/note="MSTC repeat: matches 97. .1 of consensus"
repeat_region complement(13870. .15565)
/note="MST-INTERNAL repeat: matches 1651. .1 of consensus"
repeat_region complement(15561. .15601)
/note="MSTD repeat: matches 382. .342 of consensus"
repeat_region complement(15679. .15986)
/note="MSTD repeat: matches 322. .1 of consensus"
repeat_region complement(15987. .16660)
/note="L1MC1 repeat: matches 676. .5 of consensus"
repeat_region complement(16516. .19215)
/note="L1 repeat: matches 5390. .2714 of consensus"
repeat_region complement(19218. .19320)
/note="L1MB5 repeat: matches 573. .471 of consensus"
repeat_region complement(19327. .19977)
/note="L1MB6 repeat: matches 880. .224 of consensus"
repeat_region complement(19345. .20266)
/note="L1MC3 repeat: matches 878. .1 of consensus"
repeat_region complement(20115. .20307)
/note="L1 repeat: matches 5390. .5201 of consensus"
repeat_region complement(20337. .20638)
/note="AluJb repeat: matches 302. .1 of consensus"
repeat_region complement(20661. .20882)
/note="AluJo repeat: matches 298. .79 of consensus"
repeat_region complement(20888. .21175)
/note="AluSp repeat: matches 300. .3 of consensus"
repeat_region complement(21939. .22342)
/note="MT1C repeat: matches 466. .53 of consensus"
repeat_region 22549. .22632
/note="42 copies 2 mer ca 86% conserved"
repeat_region 22552. .22635
/note="6 copies 14 mer 85% conserved"
repeat_region 22698. .23001
/note="AluSp repeat: matches 1. .303 of consensus"
prim_transcript complement(23021. .23185)
repeat_region /note="match: 5' EST T39234 clone 60330"
23819. .23854
/note="9 copies 4 mer ttat 100% conserved"
repeat_region complement(23857. .24307)
/note="L1PA9 repeat: matches 910. .454 of consensus"
repeat_region complement(24336. .24754)
/note="MLT2-internal repeat: matches 3425. .2966 of consensus"
repeat_region complement(24774. .25148)
/note="MT1C repeat: matches 414. .33 of consensus"
repeat_region complement(25194. .25333)
/note="MLT2-internal repeat: matches 2930. .2792 of consensus"
repeat_region complement(25338. .25639)
/note="AluX repeat: matches 302. .1 of consensus"
repeat_region complement(26193. .26543)
/note="THE1B repeat: matches 364. .1 of consensus"
repeat_region complement(26544. .27915)
/note="THE1B-INTERNAL repeat: matches 1580. .207 of consensus"
repeat_region 27868. .27939
/note="THE1B repeat: matches 224. .295 of consensus"
repeat_region complement(27972. .28199)
/note="MLT2-internal repeat: matches 2626. .2397 of consensus"

28319. .28452
/note="MIR2 repeat: matches 1. .145 of consensus"
repeat_region complement(29645. .29945)
/note="AluX repeat: matches 301. .1 of consensus"
29946. .30076
/note="MIR repeat: matches 114. .262 of consensus"
repeat_region 30607. .30881
/note="AluJo repeat: matches 3. .299 of consensus"
repeat_region 34070. .34107
/note="MIR repeat: matches 108. .145 of consensus"
mRNA complement(join(34216. .37315,41908. .42066,53889. .53993,
62223. .62379,64236. .64927))
/gene="AML1"
/note="match: 235278"
/pseudo
/evidence-not_experimental
complement(34216. .64927)
/gene="AML1"
repeat_region 36061. .36102
/note="21 copies 2 mer tt 91% conserved"
CDS complement(join(37172. .37716,42309. .42467,54290. .54394,
62624. .62780,64637. .64918))
/gene="AML1"
/note="match: 235278"
/codon_start=1
/evidence-not_experimental
/product="AML1 (acute myeloid leukaemia)"
/protein_id="CAA18856.1"
/db_xref="GI:3900883"
/db_xref="SPTREMBL:Q13761"
/translation="MRIPVDPSTSRFTPPSPAPFCGGGGKMGKSGALSAQAANGVP
GGRRPEVRSMVDVLADHAGELVTRDSPNPLCSVLPSHMRKNTLPVAFKVALGDVP
DCTVTVTMAKNDENYSAEILRNASAVMKNOVARENDLRFVGRSGRGSFTLITVFTNP
TOVATYVRAIKVTVDGPRPRRRKQLEQTKPFDGRLERLWRVTPSTSPRGS
LSVTSFSSOPPTIOGTSELNPNPDPDFPTLPTLTERSFDPQRMHYPGMSA
APPYSATPSGTSISLSVAGMPATSRFHITLPPYPGAPQNGSGPQANPSPYHLVY
GRSSGYQFSMVAGSSGGDRSPTRMLASCTSSAASVAGNLMNPSLGGSGDGEADG
SHSNPTALSTPGRMDEAVWRPY"
repeat_region complement(38672. .38726)
/note="MIR repeat: matches 133. .80 of consensus"
repeat_region 38810. .39019
/note="MIR repeat: matches 2. .243 of consensus"

Query Match 6.2%; Score 20; DB 9; Length 95241;
Best Local Similarity 100.0%; Pred. No. 7;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 236 GTAGATTTTTCATCAATA 275
|||||
Db 17969 GTAGATTTTTCATCAATA 17988

RESULT 8
LOCUS AC026486 148269 bp DNA HTG 06-MAY-2001
DEFINITION Homo sapiens clone RP11-24K7, WORKING DRAFT SEQUENCE, 18 unordered
pieces.
ACCESSION AC026486
VERSION AC026486.2 GI:9954792
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 148269)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens, clone RP11-24K7
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 148269)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,

```

Collamore, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S., Domino, M.S., Doyle, M., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., Larocque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N., McCarthy, M., McEwan, P., McCurk, A., McKernan, K., McPheeters, R., Melchior, J., Meneus, L., Mihova, T., Miranda, C., Mienga, V., Morrow, J., Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Olivier, T.M., Oliver, J., Peterson, K., Pierre, N., Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Triglio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.

Direct Submission  
Submitted (22-MAR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Aug 31, 2000 this sequence version replaced gi:7280316.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

----- Project Information

Center project name: L4489

Center clone name: 24\_K\_7

----- Summary Statistics

Sequencing vector: M13; M77815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 140190 bases at least Q40

Consensus quality: 144274 bases at least Q30

Consensus quality: 145694 bases at least Q20

Insert size: 145000; agarose-fp

Insert size: 146569; sum-of-contigs

Quality coverage: 4.6 in Q20 bases; agarose-fp

Quality coverage: 4.5 in Q20 bases; sum-of-contigs

----- NOTE: This is a 'working draft' sequence. It currently consists of 18 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 1564: contig of 1564 bp in length  
1565 1664: gap of 100 bp  
1665 20700: contig of 19036 bp in length  
20701 20800: gap of 100 bp  
20801 21849: contig of 1049 bp in length  
21850 21949: gap of 100 bp  
21950 25604: contig of 3655 bp in length  
25605 25704: gap of 100 bp  
25705 29764: contig of 4060 bp in length  
29765 29864: gap of 100 bp  
29865 35579: contig of 5715 bp in length  
35580 35679: gap of 100 bp  
35680 41577: contig of 5898 bp in length  
41578 41677: gap of 100 bp  
41678 47413: contig of 5736 bp in length  
47414 47513: gap of 100 bp  
47514 54860: contig of 7347 bp in length  
54861 54960: gap of 100 bp  
54961 61589: contig of 6629 bp in length  
61590 61689: gap of 100 bp  
61690 68215: contig of 6526 bp in length

\* 58216 68315: gap of 100 bp  
\* 68316 78737: contig of 10422 bp in length  
\* 78738 78837: gap of 100 bp  
\* 78838 89917: contig of 11080 bp in length  
\* 89918 90017: gap of 100 bp  
\* 90018 100805: contig of 10788 bp in length  
\* 100806 100905: gap of 100 bp  
\* 100906 114689: contig of 13784 bp in length  
\* 114690 114790: gap of 100 bp  
\* 114790 132632: contig of 17843 bp in length  
\* 132633 132733: gap of 100 bp  
\* 132733 147978: contig of 15246 bp in length  
\* 147979 148078: gap of 100 bp  
\* 148079 148269: contig of 191 bp in length.

#### FEATURES

source

1. 148269  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="RP11-24K7"

/clone\_lib="RPC1-11 Human Male BAC"  
1. 1564  
/note="assembly\_fragment"

clone\_end:SP6  
vector\_side:left"

misc\_feature

misc\_feature

1665..20700  
/note="assembly\_fragment"

20801..21849  
/note="assembly\_fragment"

21950..25604  
/note="assembly\_fragment"

25705..29764  
/note="assembly\_fragment"

29865..35579  
/note="assembly\_fragment"

35680..41577  
/note="assembly\_fragment"

41678..47413  
/note="assembly\_fragment"

47514..54860  
/note="assembly\_fragment"

54961..61589  
/note="assembly\_fragment"

61690..68215  
/note="assembly\_fragment"

68316..78737  
/note="assembly\_fragment"

78838..89917  
/note="assembly\_fragment"

90018..100805  
/note="assembly\_fragment"

100906..114689  
/note="assembly\_fragment"

114790..132632  
/note="assembly\_fragment"

132733..147978  
/note="assembly\_fragment"

148079..148269  
/note="assembly\_fragment"

clone\_end:T7  
vector\_side:right"

BASE COUNT 46902 a 26848 c 26913 g 45901 t 1705 others

#### ORIGIN

Query Match 6.2%; Score 20; DB 2; Length 148269;  
Best Local Similarity 100.0%; Pred. No. 7;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 86 TGAAGGAGCCTTTGGAAATC 105

|||||

DB 48091 TGAAGGAGCCTTTGGAAATC 48110

## RESULT 9

AC083922/c  
 LOCUS AC083922 158312 bp DNA HTG 26-FEB-2001  
 DEFINITION Homo sapiens chromosome 8 clone RP11-63H17 map 8, WORKING DRAFT  
 SEQUENCE, 5 unordered pieces.  
 AC083922  
 VERSION AC083922.2 GI:13123241  
 KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
 SOURCE human.  
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 158312)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,

Bouckhaghter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,

Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,

DeArelano,K., Dewar,K., Diaz,J.S., Dodge,S., Ferreira,P.,

FitzHugh,W., Gage,D., Galagan,J., Gardyna,S., Ginde,S., Goyette,M.,

Graham,L., Grand-pierre,N., Hagos,B., Heaford,A., Horton,L.,

Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A., LaRoque,K.,

Lamarez,R., Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G.,

Macdonald,P., Marquis,N., McCarthy,M., McEwan,P., McKernan,K.,

McPheeters,R., Meldrim,J., Meneus,L., Mihova,T., Mienga,V.,

Morrow,J., Murphy,T., Naylor,J., Norman,C.H., O'Connor,T.,

O'Donnell,P., O'Neill,D., Oliver,T.M., Oliver,J., Peterson,K.,

Pierre,N., Pisani,C., Pollara,V., Raymond,C., Rieback,M., Riley,R.,

Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S., Severi,P.,

Sougnuez,C., Spencer,B., Stange-Thomann,N., Stojanovic,N.,

Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,

Tirrell,A., Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A.,

Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J.,

Zimmer,A. and Zody,M.

Direct Submission

Submitted (06-OCT-2000) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

On Feb 25, 2001 this sequence version replaced gi:10697459.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence.submissions@genome.wi.mit.edu

----- Project Information

Center project name: L11258

Center clone name: 63\_H\_17

----- Summary Statistics

Sequencing vector: Plasmid; n/a; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 157158 bases at least Q40

Consensus quality: 157649 bases at least Q30

Consensus quality: 157821 bases at least Q20

Insert size: 157912; sum-of-contigs

Quality coverage: 8.4 in Q20 bases; sum-of-contigs

-----

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 5 contigs. The true order of the pieces

\* is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence

\* as soon as it is available and the accession number will

\* be preserved.

\* 1 20482: contig of 20482 bp in length

\* 20483 20582: gap of 100 bp

\* 20583 34457: contig of 13875 bp in length

\* -----

## FEATURES

source

1. 158312

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="8"

/map="8"

/clone="RP11-63H17"

/clone\_lib="RPCI-11 Human Male BAC"

1. 20482

/note="assembly\_fragment"

clone\_end:SP6

vector\_side:left

20583..34457

/note="assembly\_fragment"

34558..51848

/note="assembly\_fragment"

51949..100652

/note="assembly\_fragment"

100753..158312

/note="assembly\_fragment"

clone\_end:T7

vector\_side:right

46458 a 29361 c 29496 g 52597 t 400 others

BASE COUNT

ORIGIN

Query Match 6.2%; Score 20; DB 2; Length 158312;

Best Local Similarity 100.0%; Pred. No. 7;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 204 TGTAACACACTTCAGTTCA 223

|||||

Db 113950 TGTAACACACTTCAGTTCA 113931

RESULT 10

AC020682

LOCUS AC020682 162249 bp DNA HTG 19-JUL-2000

DEFINITION Homo sapiens clone RP11-24G18, WORKING DRAFT SEQUENCE, 13 unordered

pieces.

AC020682

VERSION AC020682.3 GI:9280776

KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 162249)

Birren,B., Linton,L., Nusbaum,C. and Lander,E.

Homo sapiens chromosome, clone RP11-24G18

Unpublished

2 (bases 1 to 162249)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N.,

Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Bada,F.,

Boguslavsky,L., Bouckhaghter,B., Brown,A., Burkett,G., Castle,A.,

Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,

DeArelano,K., Dewar,K., Domino,M., Doyle,M., Feneator,J.,

Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,

Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,

Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,

Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G., Locke,K.,

Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,

McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,

Norman,C.H., O'Connor,T., O'Donnell,P., Oliver,T.M., Peterson,K.,

Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,

Roy,A., Santos,R., Severi,P., Spencer,B., Stange-Thomann,N.,

Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,

-----





database can be found at  
[http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) This sequence  
was generated from part of bacterial clone configs of human  
chromosome 10, constructed by the Sanger Centre Chromosome 10  
Mapping Group. Further information can be found at  
<http://www.sanger.ac.uk/HGP/Chr10>  
Rp11-63A2 is from the library RPII-11.1 constructed by the group of  
Pieter de Jong. For further details see  
<http://www.chori.org/bacpac/home.htm>  
VECTOR: pBAC3.6

This sequence is the entire insert of clone Rp11-63A2 The true  
right end of clone Rp11-809M12 is at 48921 in this sequence.

## FEATURES

source

Location/Qualifiers

1..166973

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="10"

/clone="Rp11-63A2"

/clone\_lib="RPII-11.1"

3..805

/note="L1MC5 repeat: matches 7127..7913 of consensus"

1250..1679

/note="MLR2B repeat: matches 1..444 of consensus"

1682..4635

/note="HERVL repeat: matches 2635..5652 of consensus"

4636..4874

/note="Alusq repeat: matches 67..297 of consensus"

4875..7529

/note="HERVL repeat: matches 1..2635 of consensus"

7530..7577

/note="MLR2A repeat: matches 406..453 of consensus"

7594..8007

/note="MLR2A repeat: matches 150..450 of consensus"

8233..8517

/note="MLR1-INTERNAL repeat: matches 932..1211 of  
consensus"

8555..8854

/note="L1MC/D repeat: matches 5313..5604 of consensus"

8913..9043

/note="L1MC/D repeat: matches 5604..5735 of consensus"

9108..9312

/note="L1MC5 repeat: matches 7273..7496 of consensus"

9469..9845

/note="L1MC5 repeat: matches 7507..7895 of consensus"

10014..10105

/note="46 copies 2 mer aa 63% conserved"

10601..10716

/note="L1M4 repeat: matches 2695..2818 of consensus"

10829..11016

/note="L1M4 repeat: matches 2957..3146 of consensus"

11333..11390

/note="MLR2D repeat: matches 1..58 of consensus"

11411..11516

/note="L1M4 repeat: matches 2957..3062 of consensus"

11612..11968

/note="MLR2D repeat: matches 70..435 of consensus"

11986..12049

/note="32 copies 2 mer aa 67% conserved"

12683..12935

/note="L1ME3A repeat: matches 5669..5936 of consensus"

13751..14155

/note="L1ME repeat: matches 471..848 of consensus"

14156..14460

/note="AluJb repeat: matches 1..300 of consensus"

14461..14976

/note="L1ME repeat: matches 848..1572 of consensus"

17233..17352

/note="3 copies 40 mer 75% conserved"

17242..17345

/note="26 copies 4 mer gaag 72% conserved"

17276..17359

/note="14 copies 6 mer agaaag 65% conserved"

18041..18273

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

repeat\_region

/note="MIR repeat: matches 8..246 of consensus"  
19550..19633  
/note="42 copies 2 mer ta 83% conserved"  
19554..19633  
/note="2 copies 40 mer 91% conserved"  
20024..22103  
/note="L1PB3 repeat: matches 4126..6150 of consensus"  
26100..26169  
/note="35 copies 2 mer at 77% conserved"  
26101..26164  
/note="16 copies 4 mer tata 79% conserved"  
26106..26165  
/note="5 copies 12 mer 81% conserved"  
26204..27088  
/note="L1MC3 repeat: matches 6716..7586 of consensus"  
27760..27918  
/note="MLT1J repeat: matches 201..346 of consensus"  
28811..28944  
/note="MIR repeat: matches 12..149 of consensus"  
29062..29149  
/note="MLT1D repeat: matches 2..89 of consensus"  
30829..30884  
/note="MIR repeat: matches 76..131 of consensus"  
33710..33949  
/note="MIR repeat: matches 13..262 of consensus"  
34018..34106  
/note="L2 repeat: matches 2655..2747 of consensus"  
34360..34452  
/note="MIR repeat: matches 50..144 of consensus"  
35653..35789  
/note="MIR repeat: matches 3..137 of consensus"  
36130..36201  
/note="12 copies 6 mer atat 69% conserved"  
36131..36202  
/note="36 copies 2 mer ta 70% conserved"  
37945..37968  
/note="12 copies 2 mer tc 100% conserved"  
38866..39253  
/note="L1MA3 repeat: matches 5912..6304 of consensus"  
39256..39337  
/note="THE1B repeat: matches 1..82 of consensus"  
39340..41088  
/note="L1M1 repeat: matches 971..2727 of consensus"  
41115..41324  
/note="L1MC5 repeat: matches 7701..7912 of consensus"  
41359..41767  
/note="L1MB8 repeat: matches 5725..6145 of consensus"  
41771..41943  
/note="L1MC5 repeat: matches 7603..7774 of consensus"  
43197..43436  
/note="MIR repeat: matches 16..260 of consensus"  
43580..43882  
/note="AlusC repeat: matches 1..305 of consensus"  
44598..44892  
/note="Alusq repeat: matches 1..296 of consensus"  
45115..45156  
/note="L1R16B repeat: matches 113..160 of consensus"  
49483..49604  
/note="MIR repeat: matches 141..262 of consensus"  
49614..49668  
/note="L2 repeat: matches 2626..2682 of consensus"  
50064..50218  
/note="MIR repeat: matches 12..197 of consensus"  
50244..50526  
/note="Alusx repeat: matches 1..286 of consensus"  
51305..51539  
/note="L1PA16 repeat: matches 5913..6157 of consensus"  
52439..52618  
/note="MER58A repeat: matches 19..224 of consensus"  
52647..52871  
/note="MIR repeat: matches 2..230 of consensus"  
54233..54312  
/note="2 copies 40 mer 95% conserved"

```
repeat_region 55178..55369
/Note="MT1A1 repeat: matches 1..180 of consensus"
repeat_region 55428..55613
/Note="MT1A1 repeat: matches 181..374 of consensus"
repeat_region 55992..56287
/Note="AluX repeat: matches 3..307 of consensus"
repeat_region 57708..57735
/Note="14 copies 2 mer aa 100% conserved"
repeat_region 59512..59581
/Note="35 copies 2 mer tc 92% conserved"
repeat_region 59514..59581
/Note="17 copies 4 mer tctc 94% conserved"
repeat_region 60185..60295
/Note="MIR repeat: matches 6..120 of consensus"
repeat_region 61009..61056
/Note="8 copies 6 mer tacaca 79% conserved"
repeat_region 61011..61054
/Note="11 copies 4 mer caca 81% conserved"
repeat_region 61281..62062
/Note="L2 repeat: matches 1849..2714 of consensus"
repeat_region 62072..62276

Query Match 6.2%; Score 20; DB 9; Length 166973;
Best Local Similarity 100.0%; Pred. No. 7;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 86 TGAAGGAGCCTTGGAAATC 105
|||||
Db 12948 TGAAGGAGCCTTGGAAATC 12967

RESULT 12
AL356952
LOCUS AL356952 188026 bp DNA HTG 21-JUL-2001
DEFINITION Homo sapiens chromosome 10 clone RP11-809M12, *** SEQUENCING IN
PROGRESS ***, in ordered pieces.
ACCESSION AL356952
VERSION AL356952.17 GI:15020905
KEYWORDS HTG; HTGS_PHASE2; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 188026)
Blakey,S.
Direct Submission
Submitted (20-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Jul 25, 2001 this sequence version replaced gi:14970342.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: ba809M12
----- Summary Statistics
Sequencing program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
Chemistry: Dye-terminator Big Dye; 97% of reads
Chemistry: Dye-primer Big Dye; 2% of reads
Consensus quality: 187552 bases at least Q40
Consensus quality: 187814 bases at least Q30
Consensus quality: 187920 bases at least Q20
Insert size: 188026; sum-of-contigs
Insert size: 167487; 12.8% error; agarose-fp
Quality coverage: 6.20x in Q20 bases; sum-of-contigs Quality
coverage: 6.90x in Q20 bases; agarose-fp
-----
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
```

```
* the accession number will be preserved.
Location/Qualifiers
1..188026
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP11-809M12"
/clone_lib="RPC1-11.3"
/misc_feature 1..188026
/Note="assembly_fragment:01381"
BASE COUNT 61435 a 34467 c 34626 g 57498 t
ORIGIN
1
Query Match 6.2%; Score 20; DB 2; Length 188026;
Best Local Similarity 100.0%; Pred. No. 7;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 86 TGAAGGAGCCTTGGAAATC 105
|||||
Db 154052 TGAAGGAGCCTTGGAAATC 154071

RESULT 13
AC011498
LOCUS AC011498 273403 bp DNA HTG 26-JUL-2001
DEFINITION Homo sapiens clone CTB-50L17, *** SEQUENCING IN PROGRESS ***, 1
ordered pieces.
ACCESSION AC011498
VERSION AC011498.6 GI:15022009
KEYWORDS HTG; HTGS_PHASE2; HTGS_ACTIVEFIN.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 273403)
DOE Joint Genome Institute and Stanford Human Genome Center.
Direct Submission
2 (bases 1 to 273403)
DOE Joint Genome Institute.
Submitted (07-OCT-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Jul 26, 2001 this sequence version replaced gi:13699566.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* been provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 273403: contig of 273403 bp in length.
----- Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 83685
Center clone name: CTB-50L17
-----
Summary Statistics
Consensus quality: 273020 bases at least Q40
Consensus quality: 273294 bases at least Q30
Consensus quality: 273384 bases at least Q20
Estimated insert size: 230000; agarose-fp estimation
Estimated insert size: 273402; sum-of-contigs
estimation
estimation
Quality coverage: 13.18 in Q20 bases; agarose-fp
estimation
```

estimation. Quality coverage: 11.09 in Q20 bases; sum-of-contigs

## FEATURES

source

BASE COUNT 64486 a 73388 c 76106 g 59423 t

ORIGIN

Query Match 6.2%; Score 20; DB 2; Length 273403;  
Best Local Similarity 100.0%; Pred. No. 7;  
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 32 CCAGAGATAGAGGTGGA 51

Db 156349 CCAGAGATAGAGGTGGA 156568

## RESULT 14

AB011796

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

gene

CDS

misc\_feature

gene

gene

CDS

misc\_feature

gene

gene

CDS

misc\_feature

gene

gene

CDS

misc\_feature

gene

gene

CDS

misc\_feature

gene

gene

CDS

misc\_feature

gene

gene

CDS

misc\_feature

gene

gene

CDS

misc\_feature

gene

gene

CDS

misc\_feature

gene

gene

CDS

misc\_feature

gene

gene

CDS

misc\_feature

gene

gene

CDS

misc\_feature

gene

gene

CDS

misc\_feature

gene

gene

CDS

misc\_feature

gene

gene

CDS

misc\_feature

gene

gene

CDS

misc\_feature

gene

gene

CDS

misc\_feature

gene

gene

CDS

misc\_feature

gene

gene

CDS

misc\_feature

gene

gene

CDS

misc\_feature

gene

gene

CDS

misc\_feature

gene

gene

CDS

misc\_feature

gene

gene

CDS

misc\_feature

gene

gene

CDS

misc\_feature

gene

gene

CDS

misc\_feature

gene

gene

CDS

misc\_feature

gene

gene

CDS

misc\_feature

gene

gene

CDS

misc\_feature

gene

gene

CDS

misc\_feature

gene

gene

CDS

misc\_feature

gene

gene

CDS

misc\_feature

gene

gene

CDS

misc\_feature

gene

gene

CDS

misc\_feature

gene

gene

CDS

misc\_feature

gene

gene

CDS

misc\_feature

gene

gene

CDS

misc\_feature

gene

gene

CDS

misc\_feature

gene

gene

CDS

misc\_feature

gene

gene

CDS

misc\_feature

gene

gene

CDS

misc\_feature

gene

gene

CDS

misc\_feature

gene

gene

CDS

misc\_feature

gene

gene

CDS

misc\_feature

gene

gene

CDS

misc\_feature

gene

gene

CDS

misc\_feature

gene

gene

CDS

misc\_feature

gene

gene

CDS

misc\_feature

gene

gene

CDS

misc\_feature

gene

gene

CDS

misc\_feature

gene

gene

CDS

misc\_feature

gene

gene

CDS

misc\_feature

gene

gene

CDS

misc\_feature

gene

gene

CDS

misc\_feature

gene

gene

CDS

misc\_feature

gene

gene

CDS

misc\_feature

gene

gene

CDS

misc\_feature

gene

gene

CDS

misc\_feature

gene

gene

CDS

misc\_feature

gene

gene

CDS

misc\_feature

gene

gene

CDS

misc\_feature

gene

gene

CDS

misc\_feature

gene

gene

CDS

misc\_feature

gene

gene

CDS

misc\_feature

gene

gene

CDS

misc\_feature

gene

gene

CDS

misc\_feature

gene

gene

CDS

misc\_feature

gene

gene

CDS

misc\_feature

gene

gene

CDS

misc\_feature

gene

gene

CDS

misc\_feature

gene

gene

CDS

misc\_feature

gene

```

/gene
CDS
/codon_start=1
/product="vif"
/protein_id="AAF80533.1"
/db_xref="GI:8886634"
/translation="MENRWQVIMVQVARMIRTWNSLVKHIMTVSKKAKGWLXHHY
ESRHPVSVSEVHVIPIEDMLVITTYWGLHTGERWHLGOGVYSIEWRRKRYRTQVDPPEL
ADOLIHMHYDFCPFAESAIRKAILGHRVSPCEYPAGRNKVGSLQYLALCALVTPKRTK
PPLPSVRKLTEDRNKPKQTRGRGNHTVNC"
4861..5151
/gene="vpr"
CDS
4861..5151
/gene="vpr"
/note="vpr protein"
/codon_start=1
/product="vpr"
/protein_id="AAF80534.1"
/db_xref="GI:8886635"
/translation="MEQAPEDQGPQREPYSEWMLELLEELKEAVRHFPRLWHLGLO
HIYNTYGDTEWEGVEAIRILQQLLFVHFRICQHSRIGIIRGRGRNCARRY"
5132..7742
/gene="tat"
CDS
join(5132..5346,7652..7742)
/gene="tat"
/note="tat protein"
/codon_start=1
/product="tat"
/protein_id="AAF80535.1"
/db_xref="GI:8886636"
/translation="MEPGDINIDPWNQPSQPKTACQCCKRCYHCQHCFLKKGLG
ISYGRKRWQRRTASRSSKNHQDPIPEQPLQASRNPTGPEPKVESKAEPAPFD"
5271..7926
/gene="rev"
CDS
join(5271..5346,7652..7926)
/gene="rev"
/note="rev protein"
/codon_start=1
/product="rev"
/protein_id="AAF80536.1"
/db_xref="GI:8886637"
/translation="MAGRSGSDAQLLAARIKIILYQSSPYPKPAGTRQARRNQRRR
WRARQNLHSIQRIILSTCLGRSEBPVPLPPIERLHLDSEDCGNSGTQGVGESQI
PGESSTVLGPRTTE"
5534..8068
/gene="env"
CDS
5534..8068
/gene="env"
/note="env polyprotein"
/codon_start=1
/product="env"
/protein_id="AAF80537.1"
/db_xref="GI:8886638"
/translation="MRVKETQRNHNINRWGLMIFGMLMNICNAEKMWVTVYGVVWVR
DADTILFCASDARTYSTEKHNWATHSCVPTDNPRLSENVTENFMNKKNNVEQM
HEDVILSWDESLKPCVNLITPLCVALNCTDARDNATDKNCTFNITTELEDKKNNERL
FYRLDVVPIINDSSNIGOYRLINCNSAVKQACSKVTFEPIPIHYCAPAGFAILKCRD
KEFNGTQCKNVSTVQCTHGIKPVSTOLLNGLSAGEVRIENSELNNAKTIIVQL
NSTVIRNCTRNNTKGIHIGPQAFYATGDIIGDKQAHCVSRADWNNTLQQVAE
QLHNFKNTIVFNHSGDLEVTTHSFNCGEFFYCNSTGLFNSTWYANDNTSTQDM
QSNPTITLPCIKQIINMQVQAVYAPPIQGLRCDNSITGLLTRDGGEGNDTIG
KETFRPTGDMNRNWRSELYKYKVEIKPLGVAPTKAKRRVVEREKRAVGIGVLLGF
LGAAGSTMGAASITLTVOAROLLSCIVQOONLLKATVQOHLKLTVMGIKLOQARV
LAVERYLKDQOLLGWCSCGLICTTVPNVNSKNKYINWEINENWTWIKWREID
NYTITITLIEESQNEQNEQDLALDKWASLWNFDISNLWYIRIFIMIVGGLIG
LRIIFVLISVNRQGYSPISQALTPSQEPDPGPTKEGGEQRTSSIRLVSGF
LALAMDDLRLNCLFLYHQLRDFILIVARIVETLGRGWESLKYLGKLNLAQYWGQELKSS
AISLLNATAIAVEGTDRIIEVAHRAURLINIPRIRQGFERALL"
8070..8693
/gene="nef"
CDS
8070..8693
/gene="nef"
/note="nef protein"
/codon_start=1
/product="nef"

```

```

/protein_id="AAF80538.1"
/db_xref="GI:8886639"
/translation="MGGNWSKSSIVGVWPEIRERLRRTPTPTAAAGVGAASKDLERHGA
ITSSNTSQNNAACAWLEAQEDEGVFPVRPQVPLRPMYKGAFLDSFFLKEGGLDGL
IYSKRQETILDVHNTQGFDPDMQNTPGGIRYPLCGMCKVLPVDPREVVEANE
GENCLLHPMSOHGTEDEGDEVLRWKFDSSSLARRHIARELHPDFYKD"
BASE COUNT 3248 a 1564 c 2178 g 1945 t
ORIGIN
Query Match 5.9%; Score 19; DB 14; Length 8935;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 111 CTTTCATGTCCTTCTGGAG 129
|||||
DB 3572 CTTTCATGTCCTTCTGGAG 3554
Search completed: March 30, 2002, 07:17:31
Job time: 7243 sec

```



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2002, 04:19:22 ; Search time 1557.52 Seconds  
(without alignments)  
2207.772 Million cell updates/sec

Title: US-09-095-478A-4  
Perfect score: 320  
Sequence: 1 GAAATAATGTGTAATGTTAT.....ATTACAGGACCCCTCCTTGT 320

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: em\_estfun:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estom:\*  
5: em\_estpl:\*  
6: em\_estba:\*  
7: em\_estro:\*  
8: em\_estov:\*  
9: em\_hic:\*  
10: gb\_estl:\*  
11: gb\_est2:\*  
12: gb\_hic:\*  
13: gb\_gss:\*  
14: em\_gss\_fun:\*  
15: em\_gss\_hum:\*  
16: em\_gss\_inv:\*  
17: em\_gss\_pln:\*  
18: em\_gss\_pro:\*  
19: em\_gss\_rod:\*  
20: em\_gss\_vrt:\*  
21: em\_gss\_other:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	269	84.1	1842	12 AK017003	AK017003 Mus muscu
2	107.4	33.6	544	10 AF247150	AF247150 AF247150
3	79.6	24.9	934	13 AO744125	AO744125 HS_5507_A
4	55.4	17.3	500	10 AW630609	AW630609 hb85a11.y
5	55.4	17.3	583	10 AW607690	AW607690 MR3-HT048
6	48	15.0	442	11 BF835032	BF835032 RC3-HT086
7	47.2	14.8	545	10 BE693376	BE693376 RC1-BT031
8	46.6	14.6	629	10 AW822729	AW822729 uq17g09.y
9	46	14.4	413	11 BF832095	BF832095 PM3-HT090
10	46	14.4	966	11 BG167247	BG167247 602342507
11	45.8	14.3	2903	12 AF077031	AF077031 Homo sapi
12	44.4	13.9	452	10 AW820088	AW820088 QVO-ST029

13	44.2	13.8	475	10 AW504119	AW504119 UI-HF-BNO
14	43.8	13.7	296	10 BE763750	BE763750 KC5-NT005
15	43.6	13.6	560	10 BE223330	BE223330 kp81h08.y
16	43.6	13.6	564	10 BE580815	BE580815 kp81h08.y
17	42.8	13.4	598	10 AV616209	AV616209 AV616209
18	42.6	13.3	493	10 AI396994	AI396994 fb08b12.y
19	42.2	13.2	475	10 AW839147	AW839147 CMO-LT006
20	41.8	13.1	701	11 BG484180	BG484180 602504794
21	41.6	13.0	355	11 BE830796	BE830796 CMA-MT002
22	41.2	12.9	479	11 BF775055	BF775055 285103 MA
23	40.2	12.6	300	10 AW890822	AW890822 RC5-NT005
24	40.2	12.6	571	10 AW839966	AW839966 MR4-LT007
25	40	12.5	649	11 EG829296	EG829296 602753004
26	39.6	12.4	310	10 BE395495	BE395495 601309832
27	39.2	12.3	495	10 AW503474	AW503474 UI-HF-BNO
28	39.2	12.3	634	10 AW822725	AW822725 uq17g02.y
29	38.8	12.1	584	10 AI155087	AI155087 ud59a10.y
30	38.2	11.9	376	10 AA728886	AA728886 nv37e08.f
31	38.2	11.9	444	13 AQ333088	AQ333088 HS_5010_A
32	38.2	11.9	679	13 AQ385814	AQ385814 RPK111-15
33	38.2	11.9	876	13 AZ674685	AZ674685 ENTPEH28TR
34	38.2	11.9	927	13 AZ675096	AZ675096 ENTPEH28TR
35	38	11.9	591	11 BG793262	BG793262 UTSW SM10
36	38	11.9	622	10 AW822665	AW822665 uq14b07.y
37	37.6	11.8	720	11 BF030754	BF030754 601557471
38	37.4	11.7	796	10 AO132052	AO132052 AO132052
39	37.4	11.7	811	10 AO132040	AO132040 AO132040
40	37.2	11.6	545	10 AW877282	AW877282 MR4-PT005
41	37.2	11.6	680	13 CWS02C2V	AW190624 Tetradon
42	37	11.6	411	11 BG726429	BG726429 sad44f01.
43	37	11.6	711	10 AI827034	AI827034 w107g04.x
44	37	11.6	926	11 BE799090	BE799090 601588508
45	36.8	11.5	736	11 BE849002	BE849002 uw01g10.y

## ALIGNMENTS

## RESULT : 1

AK017003 1842 bp mRNA 05-JUL-2001  
LOCUS Mus musculus adult male testis cDNA, RIKEN full-length enriched library, clone:4933431F18, full insert sequence.  
DEFINITION  
ACCESSION AK017003  
VERSION AK017003.1 GI:128556041  
KEYWORDS CAP trapper.  
SOURCE Mus musculus (strain:C57BL/6J) adult male testis cDNA to mRNA, clone:lib:RIKEN full-length enriched mouse cDNA library  
clone:4933431F18.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 1842)  
AUTHORS Carninci,P. and Hayashizaki,Y.  
TITLE High-efficiency full-length cDNA cloning  
JOURNAL Methods in enzymology. 303, 19-44 (1999)  
MEDLINE 99279253  
PUBMED 10349636  
REFERENCE 2 (bases 1 to 1842)  
AUTHORS Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.  
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes  
JOURNAL Genome research. 10 (10), 1617-1630 (2000)  
MEDLINE 20499374  
PUBMED 11042159  
REFERENCE 3 (bases 1 to 1842)  
AUTHORS Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsuai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yanamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,

Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.  
RIKEN integrated sequence analysis (RISA) system-384-format  
sequencing pipeline with 384 multicapillary sequencer  
Genome research. 10 (11), 1757-1771 (2000)  
20530913  
PUBMED 11076861

4 (bases 1 to 1842)  
THE RIKEN Genome Exploration Research Group Phase II Team and the  
FANTOM Consortium.  
Functional annotation of a full-length mouse cDNA collection  
Nature 409, 685-690 (2001)

5 (bases 1 to 1842)  
Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,  
Arakawa,T., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M.,  
Hanaoka,T., Hara,A., Hayatsu,N., Hiramoto,K., Hiraoka,T., Horii,F.,  
Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kato,H., Kawai,J.,  
Kojima,Y., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T.,  
Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y.,  
Okido,T., Owa,C., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H.,  
Sasaki,D., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,  
Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,  
Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yasunishi,A.,  
Yoshida,K., Yoshino,M., Muramatsu,M. and Hayashizaki,Y.  
Direct Submission

TITLE Submitted (10-JUL-2000) Yoshihide Hayashizaki, The Institute of  
Physical and Chemical Research (RIKEN), Laboratory for Genome  
Exploration Research Group, RIKEN Genomic Sciences Center (GSC),  
RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,  
Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,  
URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,  
Fax:81-45-503-9216)

COMMENT Please visit our web site (<http://genome.gsc.riken.go.jp/>) for  
further details.

cDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues. First strand cDNA was primed with a primer  
[5' GAGAGAAGAGGTCCAAAGCTCTTTTTTTTNN 3'], cDNA was  
prepared by using trehalose thermo-activated reverse transcriptase  
and subsequently enriched for full-length by cap-trapper. cDNA went  
through one round of normalization to Rot = 10.0 and subtraction to  
Rot = 100.0. Second strand cDNA was prepared with the primer  
adaptor of sequence [5'  
GAGAGAAGATCTCGAGTTAATAATTAATCCCCCCC 3']. cDNA was cleaved  
with BamHI and XhoI. cDNA of size comprised between 0.5 and 3 kb  
was selected before cloning. Vector: a modified pBluescript KS(+)  
after bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI;  
3' end: BamHI. Host: DH10B.

FEATURES  
Location/Qualifiers  
1..1842  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/db\_xref="MGD:MGI:1196295"  
/db\_xref="MGD:MGI:1903803"  
/clone="4933431R18"  
/sex="male"  
/tissue\_type="testis"  
/clone\_lib="RIKEN full-length enriched mouse cDNA library"

BASE COUNT 514 a 386 C 441 g 501 t

ORIGIN

Query Match 84.1%; Score 269; DB 12; Length 1842;  
Best Local Similarity 91.9%; Pred. No. 2.6e-68;  
Matches 295; Conservative 0; Mismatches 25; Indels 1; Gaps 1;

QY 1 GAAATAATTGTTATTCCTATCATACCAAGAGATAGAGTGGAAGTATCAAG 60  
|||||  
Db 1032 GAAAATAATTGTTATTCCTATCATACCAAGAGATAGAGTGGAGTATCAAG 1091  
|||||







[illegible]

Query Match	14.8%;	Score 47.2;	DB 10;	Length 545;
Best Local Similarity	51.1%;	Pred. No. 0.0022;		



Average insert size 1.3 kb. Library enriched for full-length clones and constructed by Life technologies. Note: this is a NIH\_MGC Library.

BASE COUNT 283 a 226 c 232 g 225 t

Query Match 14.4%; Score 46; DB 11; Length 966; Best Local Similarity 50.6%; Pred. No. 0.0056; Matches 162; Conservative 0; Mismatches 155; Indels 3; Gaps 2;

QY 1 GAAATAATTTGTAATGTTATGCTATGATAACACAGAGATAGAAGGTGAGCTTATCAAG 60  
DB 151 GAGCAAAATCCACAGTGTAGCCATGATCTCAGAAGTAGAGAGAAAAATCAAA 210  
QY 61 TGTTCAGTTACTGCCCCGTTTCTCTG--AAGGAGCCCTTTGGAATTCAAACACATTTTCATG 118  
DB 211 TGCCAGCGCTATGCCCCAACATCTAGGCAAAACAAATGCTCAGCAACAGACTTCGAG 270  
QY 119 TCCTTCTGGAGAACATTCAGATAACATCAGTATTTTGTATCCGATATNTTCAAAATGTGA 178  
DB 271 TGGCTCTGTGTAGATGTCAG-CAGCTGAAGGGCTTTGTGTGAGGCAATGACCCCTTGAAG 329  
QY 179 AGAAGTCCACAGAGAAAGAGTCACTCTGTAACACACTTGCAGTTTCATCAATGCCAGACC 238  
DB 330 ATATTACAGACAGAGGTGCGCCATATTTCTCATCTGNAATTCACCTGCGCCAGACC 389  
QY 239 ATGGCACTCCCTCCCTCAGTATGATTTTTCATCAAAATATGTCGTTATGTGAGAGAGCC 298  
DB 390 ATGATACACCTTCTCAACAGATGATGCTTACTTTATCTCTACATACAGACATCC 449  
QY 299 ACATTACAGGACCCCTCTT 318  
DB 450 ACAGATCAGGCCCAATCAAT 469

RESULT 11  
AF077031 2903 bp mRNA HTC 22-MAY-2001  
LOCUS Homo sapiens protein tyrosine phosphatase homolog mRNA, complete cds.  
DEFINITION AF077031.1 GI:4689109  
ACCESSION AF077031  
VERSION AF077031.1  
KEYWORDS human.  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
AUTHORS Liu, T., Zhang, J., Fu, G., Zhang, Q., Ye, M., Zhou, J., Wu, J., Shen, Y., Yu, M., Chen, S., Mao, M. and Chen, Z.  
TITLE Human protein tyrosine phosphatase (70zpep) homolog  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 2903)  
AUTHORS Liu, T.  
TITLE Direct Submission  
JOURNAL Submitted (10-JUL-1998) Shanghai Institute of Hematology, Shanghai  
Second Medical University, Rui-Jin Hospital, 197 Rui-Jin Road II, Shanghai 200025, P.R. China  
LOCATION/Qualifiers  
FEATURES  
source 1. .2903  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/cell\_line="NB4"  
73. .2496  
/codon\_start=1  
/product="protein tyrosine phosphatase homolog"  
/protein\_id="AAD2764.1"  
/db\_xref="GI:4689110"

TRANSLATION="MDQREILQKFLDEAQSKKTKKEEFANEFLLKLRQSKYKADKTY  
QTVAERPKNKKRYKIDLPYDSRVLESLITSDSSSYINANFKVGVGPKAYIAT  
PGPLSTLLDPWRIWEYSVLIIGMACMEYEMGKKCEYWAEPVEMOLERGFPSVC  
EAEKRSYDIITRLKVFENSETRIIYQFKNWPDHDPSSIDILELWDRYQED

DSVPICIHCSAGCGRTGVCADITYTWMLLKGIIPENSVFSLIREMRTQPSLVQTO  
EYELVYNVLELFRKQMDVIRDKHSESAKHCIPEKNHTLQADSYSPNPKSTTK  
AAKMNQQTMEIKESSDFRTSEISAKELVLHPAKSTSPDFLELNTYSPDKNAD  
TTMKWQKAFPIVGPLOKHOSLDFECSCNSKPVNAAGRVFENSKVPTTRKST  
PFELIORETKEVDSKENSFVLESOPHDSCEVEMOAKVMHVSSEALAYSLPYDSKH  
IRNASNVKHDSSALGVYSYIPLVNPYFSSWPPSGTSSKSLDLPEKQDGTFFSSLL  
LPTSTSLFYSINSHSLSPNISLNOESAVLATAPRIDDEIPPLPVWTPES  
FIVVEEGEFSPNPKSLSSAVKXIGTSLMGSTSEPKKDDSVILRPSKSLRSP  
KSEHQDRSSPPPLPERTLEFFLADDECHQAOISITYSTSYPTMENSYSKQTLK  
TSKSFTRSKSLKILRNMKKSCINCPNKPAAESVQSNSSSFLNFGFANRPSKPKGP  
RNPPPTWNI"

BASE COUNT 955 a 580 c 528 g 840 t

Query Match 14.3%; Score 45.8; DB 12; Length 2903; Best Local Similarity 50.2%; Pred. No. 0.0085; Matches 144; Conservative 0; Mismatches 137; Indels 6; Gaps 1;

QY 1 GAAATAATTTGTAATGTTATGCTATGATAACACAGAGATAGAAGGTGAGCTTATCAAG 60  
DB 427 GAATATAGTGTCTTATCATTTGATGGCATGCATGGAGTATGAAATGGCAAGAAAAAG 486  
QY 61 TGTTCAGTTACTGCCCCGTTTCTCTGAGGAGCCCTTTGGAATTCAAACACATTTTCATGTC 120  
DB 487 TGTGAGCGCTACTGGCTGAGCCAGTAGAGATGAGTGGAAATTTGGCCCTTTCTCTGTA 546  
QY 121 CTTCTGGAGAACTTTTCAGATAAATTTTGTCTCATCCGAATATTTCAAAATTTGTGAAG 180  
DB 547 TCCTGTGAGAGTGAAGAAAGAAATCTGATATATATATCAGGACTCTAAAGTT----- 600  
QY 181 AAGTCCACAGGAAAGAGTCACTCTGTAAACACTTGCAGTTTCATCAATATGGCCAGACCAT 240  
DB 601 AAGTTCAATAGTGAACACTCGAATCTATCTACCACTTTTCAATACAAAGAAATTTGGCCAGACCAT 660  
QY 241 GGCACCTCTCGCTCAGTAGATTTTTCATCAAAATATGTCGTTATGCT 287  
DB 661 GATGTACCTTCATCATATAGACCCCTATTTCTTGAGCTCATCTGGATGT 707

RESULT 12  
AW820088/c 452 bp mRNA EST 17-MAY-2000  
LOCUS QV0-ST0294-070400-186-ell ST0294 Homo sapiens cDNA, mRNA sequence.  
DEFINITION AW820088  
ACCESSION AW820088.1 GI:7913082  
VERSION  
KEYWORDS EST.  
SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 452)  
AUTHORS Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M. R., Nagai, M. A., da Silva, W. Jr., Zago, M. A., Bordin, S., Costa, F. F., Goldmann, G. H., Carvalho, A. F., Matsukuma, A., Baia, G. S., Simpson, D. H., Brunstein, A., deoliveira, P. S., Bucher, P., Jongeneel, C. V., O'Hare, M. J., Soares, F., Brentani, R. R., Reis, L. F., de Souza, S. J. and Simpson, A. J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
MEDLINE 20202663  
COMMENT Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?cl=4t2-QV0-ST0294-070)



Email: asimpon@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome project. This entry can be seen in the following URL  
(<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=6t2-RC5-NT0053-140>)  
600-022-D02&t3=2000-06-14&t4=1)  
Seq primer: puc 18 forward  
High quality sequence start: 14  
High quality sequence stop: 296.

**FEATURES**  
**source**

BASE COUNT	78 a	82 c	68 g	68 t
ORIGIN				

Query Match	13.7%	Score 43.8	DB 10	Length 296
Best Local Similarity	49.4%	Pred. No. 0.018		
Matches 114	Conservative 0	Mismatches 117	Indels 0	Gaps 0
QY	88	AAGGAGCGCTTTGGAATTCAAACACTTTTCATGTCTCTCTGGAGAACTTTTCAGATAAATCAG	147	
Db	32	AAAAGACAATGTCACGACACAGACTGCAGCTGGCTCTTGTGAGAAATGCAGCAGCTGAAG	91	
QY	148	TATTTTGTCTATCCGAATATTTTCAAATTTGTGAAGAATGCCACAGAAAGATCTCATCTGTGA	207	
Db	92	GGCTTTTGTGTGAGGCAATATGACCCCTTGAAGATATTTCAGACCAGAGAGGTGCGGCATATT	151	
QY	208	AAACACTTGGCAGTTTCATCAATGGCCAGACCATTTGGCCACTCTCTGCCTCAGTAGTATTTTTC	267	
Db	152	TCTCATCTGAATTTCACTGGCTGGCCAGACCATGATACACTTCTCAACCAGATGATCTG	211	
QY	268	ATCAAAATATGTCCGGTTATGTGAGGAGAGGCCACATTTACAGAGACCCCTCCCT	318	
Db	212	CTTACTTTTATCTCCACATGAGACACATCCACAGATCAGGCCCAATCAAT	262	

RESULT	15
BE223530	
LOCUS	560 bp mRNA EST 09-MAY-2001
DEFINITION	kp81h08.y1 TB95TW-SFHF Strongyloides stercoralis cDNA 5' similar to TR:Q15718 Q15718 PTPSIGMA PRECURSOR ; , mRNA sequence.
ACCESSION	BE223530
VERSION	BE223530.1 GI:8928766
KEYWORDS	EST.
SOURCE	Strongyloides stercoralis.
ORGANISM	Strongyloides stercoralis
	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
	Pangrolaimoidea; Strongyloididae; Strongyloides.
REFERENCE	1 (bases 1 to 560)
AUTHORS	McCarter,J., Clifton,S., Chiapelli,B., Pape,D., Martin,J., Wylie,T. Dante,M., Marra,M., Hillier,L., Kucaba,T., Theisinger,B., Bowers,Y., Gibbons,M., Ritter,E., Bennett,J., Franklin,C., Tsagarelshvili,R., Ronko,I., Kennedy,S., Maguire,L., Beck,C., Underwood,K., Steptoe, M., Allen,M., Person,B., Swaller,T., Harvey,N., Schurk,R., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Willson,R.
TITLE	The Washington Univ. Nematode EST Project, 1999
JOURNAL	Unpublished (1999)
COMMENT	Contact: McCarter JP The Washington Univ. Nematode EST Project, 1999 Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA Tel: 314 286 1800

Fax: 314 286 1810  
Email: [est@watson.wustl.edu](mailto:est@watson.wustl.edu)  
The library was constructed by Dr. Thomas Nutman and colleagues of NIAID, NIH ([nutman@nih.gov](mailto:nutman@nih.gov)). DNA sequencing by Washington University Genome Sequencing Center St. Louis.  
Seq primer: -40RP from Gibco  
High quality sequence stop: 424.

FEATURES	SOURCE
1. <i>General</i>	
2. <i>Specific</i>	
3. <i>Other</i>	
4. <i>Other</i>	
5. <i>Other</i>	
6. <i>Other</i>	
7. <i>Other</i>	
8. <i>Other</i>	
9. <i>Other</i>	
10. <i>Other</i>	
11. <i>Other</i>	
12. <i>Other</i>	
13. <i>Other</i>	
14. <i>Other</i>	
15. <i>Other</i>	
16. <i>Other</i>	
17. <i>Other</i>	
18. <i>Other</i>	
19. <i>Other</i>	
20. <i>Other</i>	
21. <i>Other</i>	
22. <i>Other</i>	
23. <i>Other</i>	
24. <i>Other</i>	
25. <i>Other</i>	
26. <i>Other</i>	
27. <i>Other</i>	
28. <i>Other</i>	
29. <i>Other</i>	
30. <i>Other</i>	
31. <i>Other</i>	
32. <i>Other</i>	
33. <i>Other</i>	
34. <i>Other</i>	
35. <i>Other</i>	
36. <i>Other</i>	
37. <i>Other</i>	
38. <i>Other</i>	
39. <i>Other</i>	
40. <i>Other</i>	
41. <i>Other</i>	
42. <i>Other</i>	
43. <i>Other</i>	
44. <i>Other</i>	
45. <i>Other</i>	
46. <i>Other</i>	
47. <i>Other</i>	
48. <i>Other</i>	
49. <i>Other</i>	
50. <i>Other</i>	
51. <i>Other</i>	
52. <i>Other</i>	
53. <i>Other</i>	
54. <i>Other</i>	
55. <i>Other</i>	
56. <i>Other</i>	
57. <i>Other</i>	
58. <i>Other</i>	
59. <i>Other</i>	
60. <i>Other</i>	
61. <i>Other</i>	
62. <i>Other</i>	
63. <i>Other</i>	
64. <i>Other</i>	
65. <i>Other</i>	
66. <i>Other</i>	
67. <i>Other</i>	
68. <i>Other</i>	
69. <i>Other</i>	
70. <i>Other</i>	
71. <i>Other</i>	
72. <i>Other</i>	
73. <i>Other</i>	
74. <i>Other</i>	
75. <i>Other</i>	
76. <i>Other</i>	
77. <i>Other</i>	
78. <i>Other</i>	
79. <i>Other</i>	
80. <i>Other</i>	
81. <i>Other</i>	
82. <i>Other</i>	
83. <i>Other</i>	
84. <i>Other</i>	
85. <i>Other</i>	
86. <i>Other</i>	
87. <i>Other</i>	
88. <i>Other</i>	
89. <i>Other</i>	
90. <i>Other</i>	
91. <i>Other</i>	
92. <i>Other</i>	
93. <i>Other</i>	
94. <i>Other</i>	
95. <i>Other</i>	
96. <i>Other</i>	
97. <i>Other</i>	
98. <i>Other</i>	
99. <i>Other</i>	
100. <i>Other</i>	

BASE COUNT	218 a	63 c	98 g	181 t
ORIGIN	unamplified library is 975 bp (range, 500-1500)."			

Query Match	13.6%;	Score 43.6;	DB 10;	Length 560;
Best Local Similarity	50.0%;	Pred. No. 0.025;		
Matches 140;	Conservative 0;	Mismatches 134;	Indels 6;	Gaps 1;
QY	1	GAATAANTTGTAATGTTATTTGCTATCATATAACACAGAGATAGNAGTGGAGTTATCAAG	60	
DB				
DB	54	GAACAGAAATCATCAGTTATAGTTATGTTTAAACAAATTTAAAGAGATAGAAAAA	113	
QY	61	TGTTGCAGTTTACTGGCCGCTTTCTCTGAAGGAGCCTTTGGAAATTTCAACACACTTTTCATGTC	120	
DB				
DB	114	AGTATGANAATTTGGCCATCAGAAAAAGTTATCAATATGGAATTTTTAGTGTGTTGAACCT	173	
QY	121	CTTCTGGGAAACTTTTCAGATPAACTCAGTATTTTGTCCATCCGAATATTTCAAATTTGTGAAG	180	
DB				
DB	174	ATTGCTGA-----ATATATATGCAATATATATATTAAGAGNAATTTAAANTGAATGAT	227	
QY	181	AAGTCACACAGGAAGAGTCACTCTGTAAAAACACTTGCAGTTTCATCAAAATGCCACGACCAT	240	
DB				
DB	228	ACAAATCAGGAATAATAAAAACAATAAGACATTTTCAATATATGGAATTTGCCCAAGACAT	287	
QY	241	GGCACTCCTCGCCTCAGTAGATTTTTTTTCATCAAAATATGTCC	280	
DB				
DB	288	GGTGTACCAGAGTCAGCTGAATATGTTTTTTTGAAGATTTGTAC	327	

Search completed: March 30, 2002, 04:50:05  
Job time: 1843 sec





Result No.	Query			ID	Description	
	Score	Match	Length DB			
1	57	17.8	8040	1	US-08-596-291-1	Sequence 1, Appli
2	57	17.8	8040	3	US-09-100-804-1	Sequence 1, Appli
3	57	17.8	8043	5	PCT-US94-0943-1	Sequence 1, Appli
4	57	17.8	8119	4	US-09-290-640-45	Sequence 45, Appli
5	47.2	14.8	5117	3	US-08-854-585-1	Sequence 1, Appli
6	47.2	14.8	5117	5	PCT-US95-05512-1	Sequence 1, Appli
7	44.2	13.8	3580	4	US-09-081-345-1	Sequence 1, Appli
8	38.8	12.1	285	1	US-08-202-389-3	Sequence 3, Appli
9	36.6	11.4	1210	1	US-08-811-481-30	Sequence 30, Appli
10	36.6	11.4	1413	2	US-08-239-276-9	Sequence 9, Appli
11	36.6	11.4	1413	3	US-08-468-579B-9	Sequence 9, Appli
12	36.6	11.4	1413	3	US-08-468-577B-9	Sequence 9, Appli
13	36.6	11.4	3311	2	US-08-239-276-10	Sequence 10, Appli
14	36.6	11.4	3311	2	US-08-468-579B-10	Sequence 10, Appli
15	36.6	11.4	3311	3	US-08-468-577B-10	Sequence 10, Appli
16	34.4	10.8	2351	2	US-08-548-159-6	Sequence 6, Appli
C 17	33.8	10.6	3229	1	US-07-777-715-8	Sequence 8, Appli
C 18	33.8	10.6	3229	3	US-08-170-126-3	Sequence 3, Appli
C 19	33.8	10.6	3229	3	US-08-954-418-3	Sequence 3, Appli
C 20	33.2	10.4	5690	2	US-08-447-484-2	Sequence 2, Appli
21	33.2	10.4	5690	2	US-08-716-679-2	Sequence 2, Appli
22	33	10.3	33	4	US-09-081-345-9	Sequence 9, Appli
C 23	33	10.3	33	4	US-09-081-345-10	Sequence 10, Appli
C 24	32.8	10.2	3969	1	US-08-241-853-1	Sequence 1, Appli
25	32.8	10.2	3969	2	US-08-850-917-1	Sequence 1, Appli
C 26	32	10.0	432	4	US-09-001-944-11	Sequence 11, Appli
C 27	32	10.0	432	4	US-09-240-004A-11	Sequence 11, Appli





APPLICATION NUMBER: US/08/854,585  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/237,940  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Borun, Michael F.  
REGISTRATION NUMBER: 25,447  
REFERENCE/DOCKET NUMBER: 27866/31954  
TELEPHONE: 312-474-6300  
TELEFAX: 312-474-0448  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5117 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 350..4364  
US-08-854-585-1

Query Match 14.8%; Score 47.2; DB 3; Length 5117;  
Best Local Similarity 51.1%; Pred. No. 1.8e-05;  
Matches 140; Conservative 0; Mismatches 128; Indels 6; Gaps 1;  
QY 1 GAAATAATTGTAATGTTATGCTATGATACACAGAGATAGAGGTGGAGTTATCAAG 60  
DB 3734 GAGAAAAATGATATGCAATCATTTATGTTGCTAAATGTTGAAACAGGGAACCAAA 3793  
QY 61 TGTTCAGTACTGGCCCGTTCTCTGAAGAGCGCTTTGGAATTCACACATTTTCATGTC 120  
DB 3794 TGTGAGGAGTATTGGCCC-----TCCAGCAGGCTCAGGACTATGGAGACATAAATGTG 3847  
QY 121 CTTCGAGAACTTTCAGATAACACAGTATTTGTCATCCGAATGTTCCGAATGGACATCAACAT 180  
DB 3848 GCAATGACATCAGAAATGTTCTCCGAAATGGACATCAGAGATTTTCACAGTGAAG 3907  
QY 181 AAGTCCAGAGAAAGTCTCTGTAACACATTCGAGTTTCATCAATGGCCAGACCAT 240  
DB 3908 ATCCAGACAGTGGAGTACCCCTCTGAGACAGTTCATTCACCTCCTGGCCAGACCAC 3967  
QY 241 GGCACCTCCTCAGTAGATTTTTCATCAAT 274  
DB 3968 GGTGTTCCGACACCACTGACCTGCTCATCAACT 4001

RESULT 6  
PCT-US95-05512-1  
Sequence 1, Application PC/TUS9505512  
GENERAL INFORMATION:  
APPLICANT: Tonks, Nicholas K. and stman, Arne  
TITLE OF INVENTION: Density Enhanced Protein Tyrosine  
Phosphatase  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &  
STREET: 233 South Wacker Drive, Suite 6300  
CITY: Chicago  
STATE: Illinois  
COUNTRY: United States of America  
ZIP: 60606  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/05512

FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Borun, Michael F.  
REGISTRATION NUMBER: 25,447  
REFERENCE/DOCKET NUMBER: 27866/31954  
TELEPHONE: 312-474-6300  
TELEFAX: 312-474-0448  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5117 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 350..4364  
PCT-US95-05512-1

Query Match 14.8%; Score 47.2; DB 5; Length 5117;  
Best Local Similarity 51.1%; Pred. No. 1.8e-05;  
Matches 140; Conservative 0; Mismatches 128; Indels 6; Gaps 1;  
QY 1 GAAATAATTGTAATGTTATGCTATGATACACAGAGATAGAGGTGGAGTTATCAAG 60  
DB 3734 GAGAAAAATGATATGCAATCATTTATGTTGCTAAATGTTGAAACAGGGAACCAAA 3793  
QY 61 TGTTCAGTACTGGCCCGTTCTCTGAAGAGCGCTTTGGAATTCACACATTTTCATGTC 120  
DB 3794 TGTGAGGAGTATTGGCCC-----TCCAGCAGGCTCAGGACTATGGAGACATAAATGTG 3847  
QY 121 CTTCGAGAACTTTCAGATAACACAGTATTTGTCATCCGAATGTTCCGAATGGACATCAACAT 180  
DB 3848 GCAATGACATCAGAAATGTTCTCCGAAATGGACATCAGAGATTTTCACAGTGAAG 3907  
QY 181 AAGTCCAGAGAAAGTCTCTGTAACACATTCGAGTTTCATCAATGGCCAGACCAT 240  
DB 3908 ATCCAGACAGTGGAGTACCCCTCTGAGACAGTTCATTCACCTCCTGGCCAGACCAC 3967  
QY 241 GGCACCTCCTCAGTAGATTTTTCATCAAT 274  
DB 3968 GGTGTTCCGACACCACTGACCTGCTCATCAACT 4001

RESULT 7  
US-09-081-345-1  
Sequence 1, Application US/09081345  
Patent No. 6228641  
GENERAL INFORMATION:  
APPLICANT: Bahija Jallal  
APPLICANT: Gregory D. Plowman  
TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF  
PTP04 RELATED DISORDERS  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Lyon & Lyon  
STREET: 633 West Fifth Street  
CITY: Suite 4700  
STATE: Los Angeles  
COUNTRY: U.S.A.  
ZIP: 90071-2066  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb  
MEDIUM TYPE: storage  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: IBM P.C. DOS 5.0  
SOFTWARE: FastSeq for Windows 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/081,345  
FILING DATE: Herewith

```

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
  APPLICATION NUMBER: US 07/983,926
  FILING DATE: 01-DEC-1992
PRIOR APPLICATION DATA:
  APPLICATION NUMBER: US 07/829,141
  FILING DATE: 31-JAN-1992
PRIOR APPLICATION DATA:
  APPLICATION NUMBER: US 07/721,112
  FILING DATE: 26-JUN-1991
ATTORNEY/AGENT INFORMATION:
  NAME: Granahan, Patricia
  REGISTRATION NUMBER: 32,227
  REFERENCE/DOCKET NUMBER: BIH92-05MA
TELECOMMUNICATION INFORMATION:
  TELEPHONE: (617) 861-6240
  TELEFAX: (617) 861-9540
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
  LENGTH: 285 base pairs
  TYPE: nucleic acid
  STRANDEDNESS: double
  TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
  NAME/KEY: CDS
  LOCATION: 1..285
US-08-202-389-3

Query Match 12.1%; Score 38.8; DB 1; Length 285;
Best Local Similarity 50.0%; Pred. No. 0.0025;
Matches 97; Conservative 0; Mismatches 97; Indels 0; Gaps

QY 83 CTCCTGAAGGAGCCTTTGGAAATTCAAACACTTTTCATGTCCTCTCGGAGAACTTTCAGATAA 142
Db 20 CTCCTGAAGGAGCCTTTGGAAATTCAAACACTTTTCATGTCCTCTCGGAGAACTTTCAGATAA 142
QY 143 CTCAGTATTTTGTCAATCCGAATATTTCAAAATTTGTAAGAAGTCCACAGGAAAGAGTCACT 202
Db 80 TTCGGAATGGACCATCAGAGATTTTGTGTGAATAATATGACAGAGTAGTGAGAGTCATC 139
QY 203 CTGTAAACACACTTGGCAGTTCATCAATGCCAGACCATGGCAGTCTCTGCCCTCAGTAGATT 262
Db 140 CTCCTGGGAGTTCACCTTTCACCTCTCGGCTGACCATGGTGTCTGTACACCCAGCCGACC 199
QY 203 TTTTCATCAATAT 276
Db 200 TGCTCATCAACTTT 213

RESULT 9
US-08-811-481-30
  Sequence 30, Application US/08811481
  Patent No. 6300093
  GENERAL INFORMATION:
    APPLICANT: Kindsvogel, Wayne
    APPLICANT: Jelinek, Laura J.
    APPLICANT: Sheppard, Paul O.
    APPLICANT: Hagopian, William A.
    APPLICANT: LaGasse, James M.
    TITLE OF INVENTION: ISLET CELL ANTIGEN 1851
    NUMBER OF SEQUENCES: 34
    CORRESPONDENCE ADDRESS:
      ADDRESSEE: ZymoGenetics, Inc.
      STREET: 1201 Eastlake Avenue East
      CITY: Seattle
      STATE: WA
      COUNTRY: USA
      ZIP: 98102
    COMPUTER READABLE FORM:
      MEDIUM TYPE: Diskette
      COMPUTER: IBM Compatible

```

```

; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ For Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/811,481
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Lingenfelter, Susan
; REGISTRATION NUMBER: P-41,156
; REFERENCE/DOCKET NUMBER: 95-36
; TELEPHONE: 206-442-6675
; TELEFAX: 206-442-6678
; TELEX:
; INFORMATION FOR SEQ ID NO: 30:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1210 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-811-481-30

Query Match 11.4%; Score 36.6; DB 4; Length 1210;
Best Local Similarity 47.8%; Pred. No. 0.022;
Matches 139; Conservative 0; Mismatches 149; Indels 3; Gaps 1;

QY 10 TGTAAATGTTATGCTATGATAACCAAGAGATAGAGGTGGAGTTATCAAGTGTTCAGT 69
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 680 TGCACCGTCATCGTCATCGTCACCGCGCTGCTGGAGATGCTGCAAGCAGTGTGACCG 739
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 70 TACTGGCCGCTTCTCTGAAGGACCTTTGGAATTCACACCTTTCATGTCCTTCTGAG 129
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 740 TACTGGCCAGATGAGGTGCTCCCTTACCACGTATATGAGGTGAACCTGGTGTGCGAG 799
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 130 AACTTTTCAGATAACTCAGTATTTTGTATCCGAATATTTCAAAATTTGGAAGAGTCCACA 189
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 800 CACATCTGG---TCGAGGACTTTCTGGTGGAGCTTCTACCTGAAGAGCTGCAGACC 856
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 190 GGAAGAGTCACTCTGTAACACCTTCAGTTTCATCAAAATGGCCAGACCACTGCTCCT 249
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 857 CAGGAGACGGCGCAGCTCAGCAGTTCCTCCTCAGCTGGCGGCGAGAGGGCACACCG 916
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 250 GCCTCAGTAGATTTTTCATCAAAATATGTCGTTATGTGAGGAGAGCCAC 300
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 917 GCCTCCAGCGGCCCTGCTGGACTTCGCGAGGAGGTGAACAAGTGCTAC 967
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 10
US-08-239-276-9
; Sequence 9, Application US/08239276
; Patent No. 5840836
; GENERAL INFORMATION:
; APPLICANT: Rabin, Daniel
; TITLE OF INVENTION: PANCREATIC ISLET CELL ANTIGENS
; TITLE OF INVENTION: OBTAINED BY MOLECULAR CLONING
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sprung Kramer Schaefer & Briscoe
; STREET: 660 White Plains Road
; CITY: Tarrytown
; STATE: New York
; COUNTRY: USA
; ZIP: 10591-5144
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4 Mb storage
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: System 7.5
; SOFTWARE: WordPerfect

```

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/239,276
; FILING DATE: 05-MAY-1994
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/872,646
; FILING DATE: 08-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/715,181
; FILING DATE: 14-JUN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/441,703
; FILING DATE: 04-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/312,543
; FILING DATE: 17-FEB-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Kurt G. Briscoe
; REGISTRATION NUMBER: 33,141
; REFERENCE/DOCKET NUMBER: MDI 251.4-KGB
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (914) 332-1700
; TELEFAX: (914) 332-1844
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1413 nucleotides
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-239-276-9

Query Match 11.4%; Score 36.6; DB 2; Length 1413;
Best Local Similarity 47.8%; Pred. No. 0.023;
Matches 139; Conservative 0; Mismatches 149; Indels 3; Gaps 1;

QY 10 TGTAAATGTTATGCTATGATAACCAAGAGATAGAGGTGGAGTTATCAAGTGTTCAGT 69
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 483 TGCACCGTCATCGTCATCGTCACCGCGCTGCTGGAGATGCTGTCACGACGTTGACCGC 542
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 70 TACTGGCCGCTTCTCTGAAGGACCTTTGGAATTCACACCTTTCATGTCCTTCTGAG 129
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 543 TACTGGCCAGATGAGGTGCTCCCTTACCACGTATATGAGGTGAACCTGGTGTGCGAG 602
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 130 AACTTTTCAGATAACTCAGTATTTTGTATCCGAATATTTCAAAATTTGGAAGAGTCCACA 189
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 603 CACATCTGG---TCGAGGACTTTCTGGTGGAGCTTCTACCTGAACACCTGCAGACC 659
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 190 GGAAGAGTCACTCTGTAACACCTTCAGTTTCATCAAAATGGCCAGACCACTGCTCCT 249
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 650 CAGGAGACGGCGCAGCTCAGCAGTTCCTCCTCAGCTGGCGGCGAGAGGGCACACCG 719
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 250 GCCTCAGTAGATTTTTCATCAAAATATGTCGTTATGTGAGGAGAGCCAC 300
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 720 GCCTCCAGCGGCCCTGCTGGACTTCGCGAGGAGGTGAACAAGTGCTAC 770
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 11
US-08-468-579B-9
; Sequence 9, Application US/08468579B
; Patent No. 5981700
; GENERAL INFORMATION:
; APPLICANT: Rabin, Daniel
; TITLE OF INVENTION: PANCREATIC ISLET CELL ANTIGENS
; TITLE OF INVENTION: OBTAINED BY MOLECULAR CLONING
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sprung Kramer Schaefer & Briscoe
; STREET: 660 White Plains Road
; CITY: Tarrytown
; STATE: New York
; COUNTRY: USA
; ZIP: 10591-5144

```











GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 30, 2002, 08:16:28 ; Search time 27.53 Seconds  
(without alignments)  
328.258 Million cell updates/sec

Title: US-09-095-478A-8  
Perfect score: 122  
Sequence: 1 DFGMMWNNCNIAMITRE.....VRKSHITGPLLVHCTAGVGR 122

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 522463 seqs, 74073290 residues

Word size : 0  
Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : A\_Geneseq\_1101.\*  
1: /SID52/gcgdata/geneseq/geneseq/AA1980.DAT.\*  
2: /SID52/gcgdata/geneseq/geneseq/AA1981.DAT.\*  
3: /SID52/gcgdata/geneseq/geneseq/AA1982.DAT.\*  
4: /SID52/gcgdata/geneseq/geneseq/AA1983.DAT.\*  
5: /SID52/gcgdata/geneseq/geneseq/AA1984.DAT.\*  
6: /SID52/gcgdata/geneseq/geneseq/AA1985.DAT.\*  
7: /SID52/gcgdata/geneseq/geneseq/AA1986.DAT.\*  
8: /SID52/gcgdata/geneseq/geneseq/AA1987.DAT.\*  
9: /SID52/gcgdata/geneseq/geneseq/AA1988.DAT.\*  
10: /SID52/gcgdata/geneseq/geneseq/AA1989.DAT.\*  
11: /SID52/gcgdata/geneseq/geneseq/AA1990.DAT.\*  
12: /SID52/gcgdata/geneseq/geneseq/AA1991.DAT.\*  
13: /SID52/gcgdata/geneseq/geneseq/AA1992.DAT.\*  
14: /SID52/gcgdata/geneseq/geneseq/AA1993.DAT.\*  
15: /SID52/gcgdata/geneseq/geneseq/AA1994.DAT.\*  
16: /SID52/gcgdata/geneseq/geneseq/AA1995.DAT.\*  
17: /SID52/gcgdata/geneseq/geneseq/AA1996.DAT.\*  
18: /SID52/gcgdata/geneseq/geneseq/AA1997.DAT.\*  
19: /SID52/gcgdata/geneseq/geneseq/AA1998.DAT.\*  
20: /SID52/gcgdata/geneseq/geneseq/AA1999.DAT.\*  
21: /SID52/gcgdata/geneseq/geneseq/AA2000.DAT.\*  
22: /SID52/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	122	100.0	122	20 AAW89252	Rat PTP10. Rattus
2	23	18.9	405	20 AAW89251	Mouse PTP05 Isofor
3	23	18.9	426	20 AAW89249	Mouse PTP05. Mus
4	23	18.9	463	20 AAW89250	Mouse PTP05 isofofor
5	7	5.7	186	20 AAW97853	Wheat ATP-phosphor
6	7	5.7	243	22 AAG91426	C glutamicum prote
7	7	5.7	257	22 AAB59369	Drosophila protein
8	7	5.7	436	20 AAY37223	Protein involved i
9	7	5.7	488	20 AAW97397	A RNA-dependent am
10	7	5.7	488	20 AAW78471	Glutamyl-tRNA(Gln)
11	6	4.9	9	14 AAR33534	Peptide for treati

12	6	4.9	9	14	AAW89252	Endotoxin lipid A
13	6	4.9	9	16	AAW71780	Peptide neutralisi
14	6	4.9	9	18	AAW21627	Antibiotic potenti
15	6	4.9	9	18	AAW21597	Antibiotic potenti
16	6	4.9	9	21	AAW56909	Peptide contained
17	6	4.9	9	21	AAW56939	Peptide contained
18	6	4.9	10	22	AAW86608	Saccharomyces cere
19	6	4.9	10	22	AAW86610	Saccharomyces cere
20	6	4.9	48	17	AAW06711	Helicobacter-speci
21	6	4.9	61	22	AAW17794	Peptide #4228 enco
22	6	4.9	61	22	AAW30304	Peptide #4341 enco
23	6	4.9	61	22	AAW05441	Peptide #4123 enco
24	6	4.9	71	21	AAW53851	Human colon cancer
25	6	4.9	73	21	AAW69156	Peptide HH2040-BF0
26	6	4.9	74	19	AAW41091	Chicken matrix met
27	6	4.9	100	21	AAW02295	Human secreted pro
28	6	4.9	105	21	AAW90691	Human spectrin ple
29	6	4.9	108	19	AAW41092	Chicken matrix met
30	6	4.9	111	22	AAW17183	Peptide #3617 enco
31	6	4.9	111	22	AAW29676	Peptide #3713 enco
32	6	4.9	111	22	AAW04878	Peptide #3560 enco
33	6	4.9	117	19	AAW48090	Aeromonas caviae p
34	6	4.9	127	21	AAW58230	Feline mature GMCS
35	6	4.9	141	21	AAW57566	Arabidopsis thalia
36	6	4.9	142	21	AAW03889	Human secreted pro
37	6	4.9	144	21	AAW58229	Feline GMCSF. Fel
38	6	4.9	144	22	AAW67463	Amino acid sequenc
39	6	4.9	144	22	AAW67464	Amino acid sequenc
40	6	4.9	159	22	AAW39275	Human polypeptide
41	6	4.9	159	22	AAW89299	Human secreted pro
42	6	4.9	179	20	AAW36002	Extended human sec
43	6	4.9	179	22	AAW39714	Human polypeptide
44	6	4.9	185	19	AAW98456	H. pylori GHPO 654
45	6	4.9	193	19	AAW41090	Chicken matrix met

ALIGNMENTS

RESULT 1  
AAW89252  
ID AAW89252 standard; Protein; 122 AA.  
XX  
AC AAW89252;  
XX  
DT 10-MAR-1999 (first entry)  
DE Rat PTP10.  
XX  
KW PTP04; PTP05; PTP10; SAD; ALP; ALK-7; protein tyrosine phosphatase;  
KW type I receptor serine/threonine kinase; cancer; leukaemia; lymphoma;  
KW neurodegenerative disease; neuronal survival; Alzheimer's disease;  
KW Parkinson's disease; Huntington's disease.  
XX  
OS Rattus sp.  
XX  
PN WO9849317-A2.  
XX  
PD 05-NOV-1998.  
XX  
PF 27-APR-1998; 98WO-US08439.  
XX  
PR 23-OCT-1997; 97US-0063595.  
PR 28-APR-1997; 97US-0044428.  
PR 20-MAY-1997; 97US-0047222.  
PR 11-JUN-1997; 97US-0049477.  
PR 11-JUN-1997; 97US-0049756.  
PR 18-JUN-1997; 97US-0049914.  
XX  
(SJGE-) SUGEN INC.  
XX  
App H, Clary D, Courtneidge SA, Hui TH, Jallal B;  
PI Markby D, Onrust S, Peles E, Plowman GD;

```

XX WPI: 1999-009434/01.
DR N-PSDB; AAV81747.
XX
XX New nucleic acid encoding specific protein tyrosine phosphatases -
PT useful for identifying specific modulators for treatment and
PT prevention of cancer and neurodegenerative disease
XX
XX Claim 2; Page 160; 193pp; English.
XX
XX The present invention describes isolated, enriched or purified nucleic
CC acids encoding PTP04, SAD, PTP05, PTP10, ALP and ALK-7 proteins. The
CC present sequence represents rat PTP10. The above proteins, other than
CC ALK-7, are protein tyrosine phosphatases (PTPs) and are used to identify
CC substances that modulate their activity (i.e. agonists and antagonists,
CC including NBP) in vivo or in vitro. These substances are used to treat
CC or prevent diseases associated with abnormal signal transduction
CC pathways that involve the proteins, particularly cancer (e.g. leukaemia
CC and lymphoma), while modulators of ALK-7 (which is a type I receptor
CC serine/threonine kinase) are used to promote neuronal survival,
CC particularly for treating Alzheimer's, Parkinson's or Huntington's
CC diseases. Nucleic acid fragments of the polynucleotides encoding the
CC proteins can be used as probes to identify and clone related sequences;
CC to detect protein-encoded RNA; to generate transgenic animals and in
CC gene therapy (optionally after mutation). Ab are used to determine the
CC proteins.
XX
XX Sequence 122 AA;
SQ
Query Match 100.0%; Score 122; DB 20; Length 122;
Best Local Similarity 100.0%; Pred. No. 3.8e-125; Indels 0; Gaps 0;
Matches 122; Conservative 0; Mismatches 0;
QY 1 DFGMMWNNCNCVIAITREIEGVIKCCSWPVSLKEPLEKHFHVLLENFQITQYFVI 60
DB 1 dfgmmwennncnviamitreieggvikccswpvsleplekfhfvllefnfqtqyfv 60
QY 61 RIQIVKSTGKSHVXHLQFIKWPDHGTPTASVDFFIKYVYVRKSHITGPLLHVHCTAGV 120
DB 61 riqivkstkgtshvkhlfikwpdhgtptasvdffikyvrvrkshitgpllvhctagv 120
QY 121 GR 122
DB 121 gr 122
RESULT 2
AAW89251
ID AAW89251 standard; Protein; 405 AA.
XX
XX AAW89251;
XX
XX 10-MAR-1999 (first entry)
XX Mouse PTP05 isoform #2.
XX
XX PTP04; PTP05; PTP10; SAD; ALP; ALK-7; protein tyrosine phosphatase;
KW type I receptor serine/threonine kinase; cancer; leukaemia; lymphoma;
KW neurodegenerative disease; neuronal survival; Alzheimer's disease;
KW Parkinson's disease; Huntington's disease.
XX
XX Mus sp.
XX
XX WO9849317-A2.
XX
XX 05-NOV-1998.
XX
XX 27-APR-1998; 98WO-US08439.
XX
XX 23-OCT-1997; 97US-0063595.
XX 28-APR-1997; 97US-0044428.
XX 20-MAY-1997; 97US-0047222.

```

```

PR 11-JUN-1997; 97US-0049477.
PR 11-JUN-1997; 97US-0049756.
PR 18-JUN-1997; 97US-0049914.
XX
XX (SUGB-) SUGEN INC.
XX
XX App H, Clary D, Courtneidge SA, Hui TH, Jallal B;
PI Markby D, Onrust S, Peles E, Plowman GD;
XX
XX WPI: 1999-009434/01.
DR N-PSDB; AAV81746.
XX
XX New nucleic acid encoding specific protein tyrosine phosphatases -
PT useful for identifying specific modulators for treatment and
PT prevention of cancer and neurodegenerative disease
XX
XX Claim 2; Page 158-160; 193pp; English.
XX
XX The present invention describes isolated, enriched or purified nucleic
CC acids encoding PTP04, SAD, PTP05, PTP10, ALP and ALK-7 proteins. The
CC present sequence represents mouse PTP05. The above proteins, other than
CC ALK-7, are protein tyrosine phosphatases (PTPs) and are used to identify
CC substances that modulate their activity (i.e. agonists and antagonists,
CC including NBP) in vivo or in vitro. These substances are used to treat
CC or prevent diseases associated with abnormal signal transduction
CC pathways that involve the proteins, particularly cancer (e.g. leukaemia
CC and lymphoma), while modulators of ALK-7 (which is a type I receptor
CC serine/threonine kinase) are used to promote neuronal survival,
CC particularly for treating Alzheimer's, Parkinson's or Huntington's
CC diseases. Nucleic acid fragments of the polynucleotides encoding the
CC proteins can be used as probes to identify and clone related sequences;
CC to detect protein-encoded RNA; to generate transgenic animals and in
CC gene therapy (optionally after mutation). Ab are used to determine the
CC proteins.
XX
XX Sequence 405 AA;
SQ
Query Match 18.9%; Score 23; DB 20; Length 405;
Best Local Similarity 100.0%; Pred. No. 9.3e-17;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 94 DFTIKYVYVRKSHITGPLLHVC 116
DB 337 dftikyrvyrkshitgpllvhc 359
RESULT 3
AAW89249
ID AAW89249 standard; Protein; 426 AA.
XX
XX AAW89249;
XX
XX 10-MAR-1999 (first entry)
XX Mouse PTP05.
XX
XX PTP04; PTP05; PTP10; SAD; ALP; ALK-7; protein tyrosine phosphatase;
KW type I receptor serine/threonine kinase; cancer; leukaemia; lymphoma;
KW neurodegenerative disease; neuronal survival; Alzheimer's disease;
KW Parkinson's disease; Huntington's disease.
XX
XX Mus sp.
XX
XX WO9849317-A2.
XX
XX 05-NOV-1998.
XX
XX 27-APR-1998; 98WO-US08439.
XX
XX 23-OCT-1997; 97US-0063595.
XX 28-APR-1997; 97US-0044428.
XX 20-MAY-1997; 97US-0047222.

```



PI Fujimori K, Mizutani M, Ohta D;  
 DR WPI: 1999-153320/13.  
 DR N-PSDB; AAX24355.  
 XX  
 PT New plant ATP-phosphoribosyl transferase (APRT) encoding DNA  
 PT useful for providing herbicide-resistant plants  
 XX  
 PS Claim 2; Page 34; 58pp; English.  
 XX  
 CC This is the amino acid sequence of an ATP phosphoribosyl transferase  
 CC (APRT) polypeptide of wheat, as encoded by a cDNA clone (see  
 CC AAX24355) isolated from a 7-day-old wheat seedling cDNA library.  
 CC Another wheat APRT polypeptide (see AAW97852) is encoded by another  
 CC isolated cDNA clone (see AAX24354). The isolation of 2 DNA sequences  
 CC coding for individual proteins is probably due to the presence of  
 CC multiple isoforms encoded by different genes in the wheat genome.  
 CC Arabidopsis APRT polypeptides (see AAW97854 and AAW97855) are also  
 CC provided. Sequence comparison suggests the wheat polypeptides  
 CC lack their N-terminal portions including a signal peptide and  
 CC chloroplast transit peptide found in the Arabidopsis APRT proteins.  
 CC Using the information obtained from these sequences, APRT enzymes  
 CC can be obtained from any plant source. The invention also  
 CC encompasses the recombinant production of APRT, probes, and methods  
 CC for detecting the presence and form of the APRT gene and  
 CC quantitating levels of APRT transcripts in an organism. Expression  
 CC cassettes and recombinant vectors are also provided, as well as the  
 CC preparation of transgenic plants, plant tissue and seed which has  
 CC been stably transformed with a recombinant DNA molecule comprising  
 CC a plant promoter and a gene encoding APRT. A DNA molecule encoding  
 CC a modified APRT, having at least 1 amino acid modification and  
 CC which is tolerant to a herbicide can be used to confer herbicide  
 CC tolerance to a plant. Recombinant APRT enzyme can be used to  
 CC screen known herbicidal chemicals to determine if they inhibit  
 CC APRT, and to screen for candidate herbicides.  
 XX  
 SQ Sequence 186 AA;  
 XX  
 Query Match 5.7%; Score 7; DB 20; Length 186;  
 Best Local Similarity 100.0%; Pred. No. 13;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 20 EIEGGVI 26  
 Db 40 EIEGGVI 46  
 |||||  
 RESULT 6  
 AAG91426  
 ID AAG91426 standard; Protein; 243 AA.  
 XX  
 AC AAG91426;  
 XX  
 DT 26-SEP-2001 (first entry)  
 XX  
 DE C glutamicum protein fragment SEQ ID NO: 5180.  
 XX  
 KW Corynebacterium; amino acid synthesis; vitamin; saccharide;  
 KW organic acid synthesis.  
 XX  
 OS Corynebacterium glutamicum.  
 XX  
 PN EP1108790-A2.  
 XX  
 PD 20-JUN-2001.  
 XX  
 PE 18-DEC-2000; 2000EP-0127688.  
 XX  
 PR 16-DEC-1999; 99JP-0377484.  
 PR 07-APR-2000; 2000JP-0159162.  
 PR 03-AUG-2000; 2000JP-0280988.  
 XX

(KYOW ) KYOWA HAKKO KOGYO KK.  
 Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;  
 Tateishi N, Senoh A, Ikeda M, Ozaki A;  
 WPI: 2001-376931/40.  
 DR N-PSDB; AAH66645.  
 XX  
 PT Novel polynucleotides derived from Coryneform bacteria, for identifying  
 PT mutation point of a gene, measuring expression of a gene, analysing  
 PT expression profile or pattern of a gene and identifying homologous gene  
 XX  
 PS Claim 17; SEQ ID NO: 5180; 246pp + Sequence Listing; English.  
 XX  
 CC The present invention provides a number of nucleotide and protein  
 CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These  
 CC are useful for identifying the mutation point of a gene derived from a  
 CC mutant of coryneform bacterium, measuring expression amount and  
 CC analysing the expression profile or expression pattern of a gene derived  
 CC from Coryneform bacterium, and identifying a homologue of a gene derived  
 CC from coryneform bacterium. Coryneform bacteria are useful for producing  
 CC amino acids, nucleic acids, vitamins, saccharides and organic acids,  
 CC particularly L-lysine. The present sequence is a protein described  
 CC in the exemplification of the invention.  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from the  
 CC European Patent Office.  
 XX  
 SQ Sequence 243 AA;  
 XX  
 Query Match 5.7%; Score 7; DB 22; Length 243;  
 Best Local Similarity 100.0%; Pred. No. 16;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 89 TPASVDF 95  
 Db 165 tpsvdf 171  
 |||||  
 RESULT 7  
 AAB59369  
 ID AAB59369 standard; Protein; 257 AA.  
 XX  
 AC AAB59369;  
 XX  
 DT 21-MAR-2001 (first entry)  
 XX  
 DE Drosophila protein tyrosine phosphatase #1.  
 XX  
 KW Protein tyrosine phosphatase; human; mouse; fruit fly; PTP;  
 KW substrate trapping.  
 XX  
 OS Drosophila sp.  
 XX  
 PN WC200075339-A1.  
 XX  
 PD 14-DEC-2000.  
 XX  
 PE 24-MAY-2000; 2000WO-US14211.  
 XX  
 PR 03-JUN-1999; 99US-0137319.  
 PR 16-JUN-1999; 99US-0334575.  
 XX  
 PA (COLD-) COLD SPRING HARBOR LAB.  
 XX  
 PI Tonks NK, Zhang S;  
 XX  
 DR WPI; 2001-080598/09.  
 XX  
 PT New substrate trapping mutant protein tyrosine phosphatases (PTP) in  
 PT which the wild type PTP catalytic domain invariant aspartate is

PT replaced with an unphosphorylated amino acid, useful in gene therapy -

PS Disclosure; Fig 1; 109pp; English.

CC The present invention provides substrate trapping mutant protein tyrosine

CC phosphatases (PTps). They can be used to reduce the activity of tyrosine

CC phosphorylated proteins and to screen for modulators capable of altering

CC the binding of protein tyrosine phosphatases to their substrate. These

CC may be used in disease diagnosis and treatment.

XX

SQ Sequence 257 AA;

Query Match 5.7%; Score 7; DB 22; Length 257;

Best Local Similarity 100.0%; Pred. No. 17;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 WPDHGTP 90

Db 156 wpdhgtp 162

RESULT 8

AAV37223

ID AAV37223 standard; Protein; 436 AA.

XX AAV37223;

AC AAV37223;

XX 07-OCT-1999 (first entry)

XX Protein involved in intermediate metabolism of nucleic acids.

XX Vaccine; eye disease; conventional trachoma; nonendemic trachoma;

KW paratrachoma; inclusion conjunctivitis; genital disease; perihhepatitis;

KW nongonococcal urethritis; epididymitis; cervicitis; salpingitis;

KW bartholinitis; pneumopathy; venereal lymphogranulomatosis.

XX Chlamydia trachomatis.

XX WO9928475-A2.

XX 10-JUN-1999.

XX 27-NOV-1998; 98WO-IB01939.

XX 04-NOV-1998; 98US-0107077.

XX 28-NOV-1997; 97FR-0015041.

XX 17-DEC-1997; 97FR-0016034.

XX (GEST ) GENSET.

XX Griffiths R;

XX WPI; 1999-371125/31.

XX Genome sequence of Chlamydia trachomatis

XX Disclosure; Page 986; 1755pp; English.

XX AAY36754-Y37949 are encoded by open reading frames (ORFs) of the genome

CC of Chlamydia trachomatis (see AA201425). The polypeptides can be used as

CC vaccines against Chlamydia trachomatis. Antisense and ribozyme sequences

CC can also be used to control growth of the microorganism. Chlamydia

CC trachomatis is responsible for a large number of diseases, e.g. eye

CC diseases such as conventional trachoma, nonendemic trachoma,

CC paratrachoma, and inclusion conjunctivitis; genital diseases such as

CC nongonococcal urethritis, epididymitis, cervicitis, salpingitis,

CC perihhepatitis, bartholinitis; pneumopathy in breast feeding infants;

CC and venereal lymphogranulomatosis. The polypeptides of the invention

CC may be of use in treating these diseases.

XX Sequence 436 AA;

Query Match 5.7%; Score 7; DB 20; Length 436;

Best Local Similarity 100.0%; Pred. No. 27;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 151 NFQITQY 57

Db 96 nfqitqy 102

RESULT 9

AAW97397

ID AAW97397 standard; Protein; 488 AA.

XX AAW97397;

XX 17-MAY-1999 (first entry)

XX A RNA-dependent amino transferase (ratB) subunit.

XX RNA-dependent amino transferase subunit; ratB; Gln-tRNA Gln;

KW protein synthesis; classic ocular trachoma; inclusion conjunctivitis;

KW genital trachoma; infant pneumonitis; Lymphogranuloma Venereum;

KW incipient trachoma; keratitis; papillary hypertrophy;

KW corneal infiltration; vulvovaginitis; ear infection; prostatitis;

KW climatic bubo; bacterial infection; sexually transmitted disease;

KW infertility; bacterial adhesion; matrix protein; wound; body implant.

XX Chlamydia trachomatis.

XX EP899339-A1.

XX 03-MAR-1999.

XX 04-AUG-1998; 98EP-0306205.

XX 13-AUG-1997; 97US-0910313.

XX (SMIK ) SMITHKLINE BEECHAM CORP.

XX Black MT;

XX WPI; 1999-144810/13.

XX N-PSDB; AAX16047.

XX New Chlamydia trachomatis RNA-dependant Amino Transferase (ratB)

subunit gene and protein - useful as diagnostic reagents and for

prevention and treatment of Chlamydia infections, which cause

sexually transmitted diseases and infant pneumonitis

XX Claim 6; Page 7; 26pp; English.

XX The present sequence represents a RNA-dependent amino transferase (ratB)

subunit polypeptide. The protein catalyses the formation of Gln-tRNA Gln,

which is required for accurate protein synthesis in prokaryotes. RatB

polypeptides and polynucleotides are useful for diagnosing diseases due

to an infection of an organism with the ratB gene, by determining the

presence of the nucleic acid encoding ratB, and/or analysing for the

presence or amount of ratB polypeptide in the sample. They can diagnose

the stage and type of infection. RatB polypeptides are also useful for

screening for compounds which affect activity of the protein. These

compounds can be used in treatment to inhibit (antagonist i.e.

antibacterial drugs) or enhance (agonist) ratB activity, in addition to

direct administration of ratB polypeptides to treat conditions associated

with a lack of ratB polypeptide, or direct administration of antisense

sequences to prevent expression. RatB polypeptides and antibodies induce

an immune response to immunise and prevent disease. Diseases diagnosed,

prevented or treated include: classic ocular trachoma, inclusion

conjunctivitis, genital trachoma, infant pneumonitis, Lymphogranuloma

Venerum, incipient trachoma, keratitis, papillary hypertrophy, corneal

infiltration, vulvovaginitis, ear infection, prostatitis, and climatic

bubo; and bacterial infections, especially Chlamydia trachomatis

infections which cause sexually transmitted diseases and infertility.

CC RatB polypeptides, polynucleotides and their (ant)agonists can prevent  
 CC adhesion of bacteria to matrix proteins, and are useful for use on wounds  
 CC and body implants to prevent bacterial infection.  
 XX  
 SQ Sequence 488 AA;

Query Match 5.7%; Score 7; DB 20; Length 488;

Best Local Similarity 100.0%; Pred. No. 30;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 51 NFOITQY 57  
 |  
 Db 94 nfgitgy 100

RESULT 10

AAW78471  
 ID AAW78471 standard; Protein: 488 AA.

XX AC AAW78471;

DT 11-MAY-1999 (first entry)

DE Glutamyl-tRNA(Gln) amidotransferase subunit ratB subunit.

XX Glutamyl-tRNA(Gln) amidotransferase; ratB; diagnosis; trachoma; adhesion;  
 KW conjunctivitis; infant pneumonitis; Lymphogranuloma Venereum; keratitis;  
 KW papillary hypertrophy; corneal infiltration; vulvovaginitis; prostaticitis;  
 KW ear infection; climatic bubo; infection; sexually transmitted disease;  
 KW infertility; antagonist; matrix protein; wound; body implant.

XX Chlamydia trachomatis.

XX EP892061-A2.

PD 20-JAN-1999.

XX 01-JUL-1998; 98EP-0305247.

XX 18-JUL-1997; 97US-0896342.

XX (SMIK ) SMITHKLINE BEECHAM CORP.

XX Black MT, Reichard R;

XX WPI; 1999-083578/08.

DR N-PSDB; AAX17999.

XX New Chlamydia trachomatis polypeptide and polynucleotide - useful  
 PT as diagnostic reagents and for prevention and treatment of diseases  
 PT and sexually transmitted diseases and infertility caused by  
 PT Chlamydia trachomatis infections

XX Claim 12; Page 6; 23pp; English.

XX This sequence represents the Chlamydia trachomatis glutamyl-tRNA(Gln)  
 CC amidotransferase subunit ratB protein sequence. RatB polypeptides and  
 CC polynucleotides are useful for diagnosing diseases e.g. classic ocular  
 CC trachoma, inclusion conjunctivitis, genital trachoma, infant pneumonitis,  
 CC Lymphogranuloma Venereum, incipient trachoma, keratitis, papillary  
 CC hypertrophy, corneal infiltration, vulvovaginitis, ear infection,  
 CC prostaticitis, and climatic bubo, and bacterial infections, especially  
 CC Chlamydia trachomatis infections which cause sexually transmitted  
 CC diseases and infertility. RatB polypeptides, polynucleotides and their  
 CC (ant)agonists can prevent adhesion of bacteria to matrix proteins, and  
 CC are useful for use on wounds and body implants to prevent bacterial  
 CC infection.

XX Sequence 488 AA;

Query Match 5.7%; Score 7; DB 20; Length 488;

Best Local Similarity 100.0%; Pred. No. 30;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 51 NFOITQY 57  
 |  
 Db 94 nfgitgy 100

RESULT 11

AAR33534  
 ID AAR33534 standard; peptide: 9 AA.

XX AC AAR33534;

DT 07-JUL-1993 (first entry)

DE Peptide for treating septic shock.

XX Toxic shock; blood endotoxin removal; serum; diagnostic reagent;  
 KW cytokine release control; treatment; pertussis; bacterial meningitis;  
 KW HIV related infections; polymyxin B.

XX Synthetic.

XX ZA9200943-A.

XX 25-NOV-1992.

PF 10-FEB-1992; 92ZA-0000943.

XX 11-FEB-1991; 91US-0658744.

XX (PORR/) PORRO M.

XX Porro M;

XX WPI; 1993-094304/11.

XX New peptide for treatment or prevention of toxic shock - comprises  
 PT specified sequences of aminoacid(s) and analogs  
 PT comprising sequences retro-orientated

XX Claim 12; Page 32; 39pp; English.

XX This peptide is a specific example of a generic peptide of  
 CC formula R-( Lys/Arg/His - Phe/Tyr/Trp - Leu/Ile/Val )n-R, where  
 CC n = 1-100 and each R is H, an amino acid residue or a fatty acid  
 CC residue. The peptide is useful for treating or preventing septic  
 CC shock, mixing with polymyxin B to reduce its toxicity; removing  
 CC endotoxins from blood, sera or other fluids (in vivo or in  
 CC vitro); controlling release of cytokines induced by endotoxins;  
 CC as diagnostic reagents to detect and quantify toxins in blood  
 CC or sera; preparing non-toxic antigenic complexes of lipid A or  
 CC lipopolysaccharide (LPS); and for treating pertussis, bacterial  
 CC meningitis and HIV-related infections. The usual dose is 10-100  
 CC ug/kg/day, given parenterally. It binds to the same sites as  
 CC polymyxin B, i.e. it inhibits all the toxic effects of lipid A. It  
 CC has no antibiotic activity; does not lyse erythrocytes; has no  
 CC toxicity in mice when injected at 50mg/kg and is relatively unstable  
 CC against proteases.

XX Sequence 9 AA;

Query Match 4.9%; Score 6; DB 14; Length 9;

Best Local Similarity 100.0%; Pred. No. 4.3e+05;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 99 YVRVVR 104  
 |  
 Db 2 yrvrvr 7



## RESULT 12

AAR39296  
ID AAR39296 standard; peptide; 9 AA.  
XX  
AC AAR39296;  
XX  
DT 22-DEC-1993 (first entry)  
XX  
DE Endotoxin lipid A neutralising peptide.  
XX  
KW Bacterial endotoxin; non-toxic; treatment; prevention; toxic shock;  
KW polymyxin B; PMB; toxicity reduction; reduce; blood endotoxin;  
KW removal; sera; vaccines; drug solns.; solutions; cytokine release;  
KW control; in vivo; in vitro; detoxification; detection;  
KW quantification.  
XX  
OS Synthetic.  
XX  
PN WO9314115-A.  
XX  
PD 22-JUL-1993.  
XX  
PF 14-MAY-1992; 92WO-EP01060.  
XX  
PR 16-JAN-1992; 92US-0819893.  
XX  
PA (PORR/) PORRO M.  
XX  
PI Porro M;  
XX  
XX WPI; 1993-243143/30.  
XX  
DR New peptide(s) which neutralise lipid A of bacterial endotoxin -  
XX forming non-toxic, antigenic complex, used to treat or prevent  
XX septic shock, in vaccines to detoxify blood, etc.  
XX  
PS Claim 12; Page 33; 45pp; English.  
XX

CC The sequence is that of a peptide which binds to the lipid A  
CC component of bacterial endotoxin at the same site as polymyxin B  
CC (PMB) and with about the same affinity to produce a non-toxic  
CC antigenic complex. Unlike PMB it is not toxic, is susceptible to  
CC proteolytic degradation in serum, has no antibiotic activity and no  
CC haemolytic action. It is especially used to treat or prevent septic  
CC shock; to reduce toxicity of PMB; to remove endotoxins from blood,  
CC sera, vaccines, drug solns. etc.; to control release of cytokines  
CC induced by endotoxins; for in vivo or in vitro detoxification of  
CC bacterial endotoxins, and to detect or quantify endotoxins in blood  
CC products.  
XX  
SQ Sequence 9 AA;

Query Match 4.9%; Score 6; DB 14; Length 9;  
Best Local Similarity 100.0%; Pred. No. 4.3e+05;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 99 YVRVVR 104  
Db 2 yvryvr 7  
|||||

## RESULT 13

AAR71780  
ID AAR71780 standard; peptide; 9 AA.  
XX  
AC AAR71780;  
XX  
DT 01-OCT-1995 (first entry)  
XX  
DE Peptide neutralising toxicity of Lipid A.  
XX  
KW Endotoxin; Lipid A; septic shock; lipopolysaccharide.

XX Synthetic.  
XX OS  
XX PN WO9503327-A.  
XX  
PD 02-FEB-1995.  
XX  
PF 21-JUL-1994; 94WO-EP02413.  
XX  
PR 26-JUL-1993; 93US-0097830.  
XX  
PA (BIOS-) BIOSYNTH SRL.  
XX  
PI Porro M;  
XX  
XX WPI; 1995-075190/10.  
XX  
XX New peptide(s) for neutralising LPS endotoxin - comprising  
XX repeating units of a basic amino acid or basic and hydrophobic  
XX amino acids  
XX  
XX Claim 11; Page 20; 26pp; English.  
XX  
XX New peptides are claimed which are linear or cyclic peptides of formula:  
XX (A)n, where A is the cationic amino acid Lys or Arg and n is 7-10;  
XX (AB)m, where A is Lys or Arg, B is the hydrophobic amino acid Val, Leu,  
XX Ile, Tyr, Phe or Trp, and m is 3 or greater; and (ABC)p, where A is  
XX Lys or Arg, B and C are Val, Leu, Ile, Tyr, Phe or Trp, and p is 2 or  
XX greater.  
XX The peptides bind to Lipid A of endotoxins and do not exhibit haemolytic  
XX activity. Hence they can be used therapeutically to treat septic shock  
XX and also in vitro to detoxify vaccines, drug solutions, injectable  
XX nutrient solutions, etc.  
XX The present sequence is a specifically claimed example of the new  
XX peptides.  
XX  
SQ Sequence 9 AA;

Query Match 4.9%; Score 6; DB 16; Length 9;  
Best Local Similarity 100.0%; Pred. No. 4.3e+05;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 99 YVRVVR 104  
Db 2 yvryvr 7  
|||||

## RESULT 14

AAW21627  
ID AAW21627 standard; peptide; 9 AA.  
XX

AC AAW21627;  
XX  
XX 26-AUG-1997 (first entry)  
XX  
XX Antibiotic potentiating peptide #39.

Potentiate; antibiotic; microbial infection; lipopolysaccharide;  
permeability; outer bacterial membrane.

XX Synthetic.

XX WO9638163-A1.

XX 05-DEC-1996.

XX 29-MAY-1996; 96WO-EP02313.

XX 31-MAY-1995; 95US-0456112.

XX (BIOS-) BIOSYNTH SRL.

PI Porro M, Varra M;  
XX WPI; 1997-034095/03.  
XX  
XX Potentiating activity of antibiotic with peptide contg. cationic  
PT amino acid sequence - reduces dose of antibiotic required  
XX  
XX Claim 43; Page 28; 37pp; English.  
PS  
XX The sequences given in AAW21589-633 represent peptides which act to  
CC potentiate the activity of an antibiotic when they are co-administered  
CC with the antibiotic. Compositions containing these peptides are used  
CC to treat or prevent microbial infections. These peptides bind to  
CC lipopolysaccharide on the bacteria so may increase permeability of  
CC the outer bacterial membrane to the antibiotic, allowing a reduction  
CC in the dose of antibiotic required by 10-90% of the normal dose for  
CC in vivo or in vitro application. Any toxic side effects are  
CC correspondingly reduced.  
XX  
XX Sequence 9 AA;  
SQ

Query Match 4.9%; Score 6; DB 18; Length 9;  
Best Local Similarity 100.0%; Pred. No. 4.3e+05;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 99 YRYVR 104  
Db | | | | |  
2 yryvr 7

RESULT 15  
AAW21597  
ID AAW21597 standard; peptide; 9 AA.  
XX  
XX AAW21597;  
XX  
XX 26-AUG-1997 (first entry)  
DE Antibiotic potentiating peptide #9.  
XX  
XX Potentiate; antibiotic; microbial infection; lipopolysaccharide;  
KW permeability; outer bacterial membrane.  
XX  
XX Synthetic.  
OS  
XX WO9638163-A1.  
PN  
XX  
XX PD 05-DEC-1996.  
XX  
XX PF 29-MAY-1996; 96WO-EP02313.  
XX  
XX PR 31-MAY-1995; 95US-0456112.  
XX  
XX (BIOS-) BIOSYNTH SRL.  
XX  
XX PI Porro M, Varra M;  
XX  
XX WPI; 1997-034095/03.  
XX  
XX Potentiating activity of antibiotic with peptide contg. cationic  
PT amino acid sequence - reduces dose of antibiotic required  
XX  
XX Claim 13; Page 24; 37pp; English.  
PS  
XX The sequences given in AAW21589-633 represent peptides which act to  
CC potentiate the activity of an antibiotic when they are co-administered  
CC with the antibiotic. Compositions containing these peptides are used  
CC to treat or prevent microbial infections. These peptides bind to  
CC lipopolysaccharide on the bacteria so may increase permeability of  
CC the outer bacterial membrane to the antibiotic, allowing a reduction  
CC in the dose of antibiotic required by 10-90% of the normal dose for  
CC in vivo or in vitro application. Any toxic side effects are

CC correspondingly reduced.  
XX  
XX Sequence 9 AA;  
SQ

Query Match 4.9%; Score 6; DB 18; Length 9;  
Best Local Similarity 100.0%; Pred. No. 4.3e+05;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 99 YRYVR 104  
Db | | | | |  
2 yryvr 7

Search completed: March 30, 2002, 08:26:44  
Job time: 616 sec



**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 30, 2002, 08:22:33 ; Search time 15.23 Seconds  
(without alignments)  
180.263 Million cell updates/sec

Title: US-09-095-478A-8  
Perfect score: 122  
Sequence: 1 DFGWMKNNCNVIAITRE.....VRKSHITGPLLVIHCTAGVGR 122

Scoring table:  
Gapop 60.0 , Gapext 60.0

Searched: 212252 seqs, 22503292 residues

Word size : 0

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Issued Patents, AA.\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7	5.7	257	2	US-08-685-992-6
2	7	5.7	257	2	US-09-144-925-6
3	7	5.7	488	4	US-08-910-313-2
4	6	4.9	9	1	US-07-819-893-10
5	6	4.9	9	1	US-08-280-397-10
6	6	4.9	9	1	US-08-097-830E-9
7	6	4.9	9	2	US-08-456-112B-9
8	6	4.9	48	2	US-08-456-112B-39
9	6	4.9	48	2	US-08-849-480A-8
10	6	4.9	105	3	US-08-407-165-3
11	6	4.9	117	2	US-08-910-856-4
12	6	4.9	260	4	US-08-957-130-15
13	6	4.9	272	2	US-08-446-345-34
14	6	4.9	327	2	US-08-907-674-3
15	6	4.9	327	2	US-09-215-087-3
16	6	4.9	327	3	US-09-391-959-3
17	6	4.9	331	2	US-08-907-674-1
18	6	4.9	331	2	US-09-215-087-1
19	6	4.9	331	3	US-09-391-959-1
20	6	4.9	440	1	US-08-061-062A-6
21	6	4.9	440	1	US-08-061-062A-8
22	6	4.9	440	3	US-08-536-150-6
23	6	4.9	440	3	US-08-536-150-8
24	6	4.9	579	4	US-08-743-168B-36
25	6	4.9	581	4	US-08-743-168B-40
26	6	4.9	590	4	US-08-743-168B-43
27	6	4.9	1174	2	US-08-446-345-36

28 6 4.9 1375 3 US-08-665-259-26 Sequence 26, Appl  
29 6 4.9 1375 3 US-08-762-500-26 Sequence 26, Appl  
30 2548 4 US-09-172-422-1 Sequence 1, Appl  
31 5 4.1 7 2 US-08-292-968-31 Sequence 31, Appl  
32 5 4.1 7 2 US-08-467-974-31 Sequence 31, Appl  
33 5 4.1 7 2 US-08-467-974-31 Sequence 31, Appl  
34 5 4.1 7 2 US-08-467-976-31 Sequence 31, Appl  
35 5 4.1 7 4 US-09-082-514-31 Sequence 31, Appl  
36 5 4.1 8 2 US-08-342-930-12 Sequence 12, Appl  
37 5 4.1 8 3 US-08-336-553A-4 Sequence 4, Appl  
38 5 4.1 8 3 US-08-336-553A-33 Sequence 33, Appl  
39 5 4.1 8 3 US-08-336-553A-55 Sequence 55, Appl  
40 5 4.1 8 3 US-08-336-553A-74 Sequence 74, Appl  
41 5 4.1 8 4 US-08-916-935-10 Sequence 10, Appl  
42 5 4.1 9 3 US-07-792-600-34 Sequence 34, Appl  
43 5 4.1 9 3 US-09-157-021-34 Sequence 34, Appl  
44 5 4.1 9 3 US-09-156-842-34 Sequence 34, Appl  
45 5 4.1 9 4 US-09-169-015-14 Sequence 14, Appl

#### ALIGNMENTS

RESULT : 1  
US-08-685-992-6  
: Sequence 6, Application US/08685992  
: Patent No. 5912138  
: GENERAL INFORMATION:  
: APPLICANT: Tonks, Nicholas  
: APPLICANT: Flint, Andrew J.  
: TITLE OF INVENTION: SUBSTRATE TRAPPING PROTEIN  
: TITLE OF INVENTION: TYROSINE PHOSPHATASES  
: NUMBER OF SEQUENCES: 36  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: HAMILTON, BROOK, SMITH & REYNOLDS, P.C.  
: STREET: Two Militia Drive  
: CITY: Lexington  
: STATE: MA  
: COUNTRY: USA  
: ZIP: 02173  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Diskette  
: COMPUTER: IBM Compatible  
: OPERATING SYSTEM: Windows 95  
: SOFTWARE: FASTSEQ for Windows Version 2.0b  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/08/685,992  
: FILING DATE: 25-JUL-1996  
: CLASSIFICATION: 435  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER:  
: FILING DATE:  
: ATTORNEY/AGENT INFORMATION:  
: NAME: Granahan, Patricia  
: REGISTRATION NUMBER: 32,227  
: REFERENCE/DOCKET NUMBER: CSHL96-03  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: 781-861-6240  
: TELEFAX: 781-861-9540  
: TELEX:  
: INFORMATION FOR SEQ ID NO: 6:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 257 amino acids  
: TYPE: amino acid  
: STRANDEDNESS: single  
: TOPOLOGY: linear  
: MOLECULE TYPE: peptide  
US-08-685-992-6

Query Match 5.7%; Score 7; DB 2; Length 257;  
Best Local Similarity 100.0%; Pred.No. 6.4;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 84 WPDHGT 90  
Db 156 WPDHGT 162

## RESULT 2

US-09-144-925-6  
; Sequence 6, Application US/09144925  
; Patent No. 5951979  
; GENERAL INFORMATION:  
; APPLICANT: Tonks, Nicholas  
; APPLICANT: Flint, Andrew J.  
; TITLE OF INVENTION: SUBSTRATE TRAPPING PROTEIN  
; TITLE OF INVENTION: TYROSINE PHOSPHATASES  
; NUMBER OF SEQUENCES: 36  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: HAMILTON, BROOK, SMITH & REYNOLDS, P.C.  
; STREET: Two Militia Drive  
; CITY: Lexington  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02421-4799  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; OPERATING SYSTEM: Windows 95  
; SOFTWARE: FastSeq for Windows Version 2.0b  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/144,925  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/685,992  
; FILING DATE: July 25, 1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Granahan, Patricia  
; REGISTRATION NUMBER: 32,227  
; REFERENCE/DOCKET NUMBER: CSHL96-03Z  
; TELEPHONE: 781-861-6240  
; TELEFAX: 781-861-9540  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 257 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-09-144-925-6

Query Match 5.7%; Score 7; DB 2; Length 257;  
Best Local Similarity 100.0%; Pred. No. 6.4;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 84 WPDHGT 90  
Db 156 WPDHGT 162

## RESULT 3

US-08-910-313-2  
; Sequence 2, Application US/08910313  
; Patent No. 6171838  
; GENERAL INFORMATION:  
; APPLICANT: Black, Michael Terence  
; TITLE OF INVENTION: No. 6171838el RatB  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dechert Price & Rhoads  
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre

; CITY: Philadelphia  
; STATE: PA  
; COUNTRY: US  
; ZIP: 19103  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/910,313  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Dickinson, Todd Q  
; REGISTRATION NUMBER: 28,354  
; REFERENCE/DOCKET NUMBER: GM10072  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 215-994-2252  
; TELEFAX: 215-994-2222  
; TELEX:  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 488 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-910-313-2

Query Match 5.7%; Score 7; DB 4; Length 488;  
Best Local Similarity 100.0%; Pred. No. 12;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 51 NFOITQY 57  
Db 94 NFOITQY 100

## RESULT 4

US-07-819-893-10  
; Sequence 10, Application US/07819893  
; Patent No. 5371186  
; GENERAL INFORMATION:  
; APPLICANT: Porro, Massimo  
; TITLE OF INVENTION: Synthetic Peptides for Detoxification  
; TITLE OF INVENTION: Of Bacterial Endotoxins and for the  
; TITLE OF INVENTION: Prevention and Treatment of Septic  
; TITLE OF INVENTION: Shock  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Hedman, Gibson, Costigan & Hoare  
; STREET: 1185 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 10036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Mb storage  
; COMPUTER: IBM PS/2  
; OPERATING SYSTEM: DOS  
; SOFTWARE: Word Perfect 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/819,893  
; FILING DATE: 19920115  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER:  
; FILING DATE:  
; ATTORNEY/AGENT INFORMATION:

Query Match 5.7%; Score 7; DB 2; Length 257;  
Best Local Similarity 100.0%; Pred. No. 6.4;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

NAME: Costigan, James V.  
REGISTRATION NUMBER: 25,669  
REFERENCE/DOCKET NUMBER: 576-002  
TELEPHONE: (212) 302-8989  
TELEFAX: (212) 302-8998  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 9 amino acids  
TYPE: AMINO ACIDS  
TOPOLOGY: linear  
US-07-819-893-10

Query Match 4.9%; Score 6; DB 1; Length 9;  
Best Local Similarity 100.0%; Pred. No. 1.6e+05;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 99 YVYVR 104  
DB 2 YVYVR 7

RESULT 5  
US-08-280-397-10  
Sequence 10, Application US/08280397  
Patent No. 5589459  
GENERAL INFORMATION:  
APPLICANT: Porro, Massimo  
TITLE OF INVENTION: Synthetic Peptides for Detoxification  
of Bacterial Endotoxins and for the  
Prevention and Treatment of Septic  
Shock  
TITLE OF INVENTION: Shock  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hedman, Gibson & Costigan, P.C.  
STREET: 1185 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Mb storage  
COMPUTER: IBM PS/2  
OPERATING SYSTEM: DOS  
SOFTWARE: Word Perfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/280,397  
FILING DATE: 07/26/94  
CLASSIFICATION: 514

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/819,893  
FILING DATE: 01/16/92  
ATTORNEY/AGENT INFORMATION:  
NAME: Costigan, James V.  
REGISTRATION NUMBER: 25,669  
REFERENCE/DOCKET NUMBER: 576-002A  
TELEPHONE: (212) 302-8989  
TELEFAX: (212) 302-8998  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 9 amino acids  
TYPE: amino acids  
TOPOLOGY: linear  
US-08-280-397-10

Query Match 4.9%; Score 6; DB 1; Length 9;  
Best Local Similarity 100.0%; Pred. No. 1.6e+05;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 99 YVYVR 104

DB 2 YVYVR 7

RESULT 6  
US-08-097-830E-9  
Sequence 9, Application US/08097830E  
Patent No. 5652211  
GENERAL INFORMATION:  
APPLICANT: Porro, Massimo  
TITLE OF INVENTION: Peptides For Neutralizing The  
Toxicity of Lipid A  
NUMBER OF SEQUENCES: 35  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hedman, Gibson & Costigan, P.C.  
STREET: 1185 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Mb storage  
COMPUTER: IBM PS/2  
OPERATING SYSTEM: DOS  
SOFTWARE: Word Perfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/097,830E  
FILING DATE:  
CLASSIFICATION: 514

PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Costigan, James V.  
REGISTRATION NUMBER: 25,669  
REFERENCE/DOCKET NUMBER: 576-003  
TELEPHONE: (212) 302-8989  
TELEFAX: (212) 302-8998  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 9 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-08-097-830E-9

Query Match 4.9%; Score 6; DB 1; Length 9;  
Best Local Similarity 100.0%; Pred. No. 1.6e+05;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 99 YVYVR 104  
DB 2 YVYVR 7

RESULT 7  
US-08-456-112B-9  
Sequence 9, Application US/08456112B  
Patent No. 5834430  
GENERAL INFORMATION:  
APPLICANT: Porro, Massimo  
TITLE OF INVENTION: POTENTIATION OF ANTIBIOTICS  
NUMBER OF SEQUENCES: 45  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hedman, Gibson & Costigan  
STREET: 1185 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Mb storage

COMPUTER: LEADING EDGE 486  
OPERATING SYSTEM: DOS  
SOFTWARE: Word Perfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/456,112B  
FILING DATE: May 31, 1995  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Costigan, James V.  
REGISTRATION NUMBER: 25,669  
REFERENCE/DOCKET NUMBER: 576-004  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 302-8989  
TELEFAX: (212) 302-8998  
INFORMATION FOR SEQ ID NO: 9:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 9 amino acids  
TYPE: amino acid  
TOPOLOGY: circular  
US-08-456-112B-9

Query Match 4.9%; Score 6; DB 2; Length 9;  
Best Local Similarity 100.0%; Pred. No. 1.6e+05;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 99 YRVYR 104  
Db 2 YRVYR 7

RESULT 8  
US-08-456-112B-39  
Sequence 39, Application US/08456112B  
Patent No. 5834430  
GENERAL INFORMATION:  
APPLICANT: Porro, Massimo  
TITLE OF INVENTION: POTENTIATION OF ANTIBIOTICS  
NUMBER OF SEQUENCES: 45  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hedman, Gibson & Costigan  
STREET: 1185 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Mb storage  
COMPUTER: LEADING EDGE 486  
OPERATING SYSTEM: DOS  
SOFTWARE: Word Perfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/456,112B  
FILING DATE: May 31, 1995  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Costigan, James V.  
REGISTRATION NUMBER: 25,669  
REFERENCE/DOCKET NUMBER: 576-004  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 302-8989  
TELEFAX: (212) 302-8998  
INFORMATION FOR SEQ ID NO: 39:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 9 amino acids  
TYPE: amino acid  
TOPOLOGY: circular

US-08-456-112B-39

Query Match 4.9%; Score 6; DB 2; Length 9;  
Best Local Similarity 100.0%; Pred. No. 1.6e+05;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 99 YRVYR 104  
Db 2 YRVYR 7

RESULT 9  
US-08-849-480A-8  
Sequence 8, Application US/08849480A  
Patent No. 5981184  
GENERAL INFORMATION:  
APPLICANT: MELCHERS, Klaus  
TITLE OF INVENTION: SCREENING MODEL  
NUMBER OF SEQUENCES: 24  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: JACOBSON, PRICE, HOLMAN & STERN, PLLC  
STREET: 400 - 7th Street, N. W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: USA  
ZIP: 20004  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/849,480A  
FILING DATE: 02-JUN-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/EP95/04711  
FILING DATE: 30-NOV-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: DE P442970.3  
FILING DATE: 02-DEC-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: DE 19505645.0  
FILING DATE: 18-FEB-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: AISENBERG, Irwin M.  
REGISTRATION NUMBER: 19,007  
REFERENCE/DOCKET NUMBER: 8125/P60984US0  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202/638-6666  
TELEFAX: 202/393-5350  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 48 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: YES  
ORIGINAL SOURCE:  
ORGANISM: Helicobacter pylori  
STRAIN: Helicobacter pylori 69A  
INDIVIDUAL ISOLATE: Clinical isolate 69A  
IMMEDIATE SOURCE:  
LIBRARY: Helicobacter pylori 69A - gene library in  
CLONE: pRH948  
US-08-849-480A-8

Query Match 4.9%; Score 6; DB 2; Length 48;  
Best Local Similarity 100.0%; Pred. No. 16;



Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 92 SVDFEI 97  
|||||  
Db 9 SVDFEI 14

RESULT 10  
US-08-407-165-3  
; Sequence 3, Application US/08407165  
; Patent No. 6054280  
; GENERAL INFORMATION:  
; APPLICANT: LEMMON, MARK A.  
; APPLICANT: FERGUSON, KATHRYN M.  
; APPLICANT: SIGLER, PAUL B.  
; APPLICANT: SCHLESSINGER, JOSEPH  
; TITLE OF INVENTION: METHODS FOR DIAGNOSIS AND TREATMENT OF  
; FILE OF INVENTION: PH DOMAIN SIGNAL TRANSDUCTION DISORDERS  
; FILE REFERENCE: 211/156  
; CURRENT APPLICATION NUMBER: US/08/407,165  
; CURRENT FILING DATE: 1995-03-20  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 3  
; LENGTH: 105  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; OTHER INFORMATION: Spectrin  
US-08-407-165-3

Query Match 4.9%; Score 6; DB 3; Length 105;  
Best Local Similarity 100.0%; Pred. No. 34;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 33 PVSKE 38  
|||||  
Db 55 PVSKE 60

RESULT 11  
US-08-910-856-4  
; Sequence 4, Application US/08910856  
; Patent No. 5981257  
; GENERAL INFORMATION:  
; APPLICANT: FUKUI, TOSHIAKI  
; APPLICANT: DOI, YOSHIHARU  
; TITLE OF INVENTION: POLYESTER SYNTHASE GENE AND PROCESS  
; FILE OF INVENTION: FOR PRODUCING POLYESTER  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: FISH & RICHARDSON P.C.  
; STREET: 4225 EXECUTIVE SQUARE, SUITE 1400  
; CITY: LA JOLLA  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 92037  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/910,856  
; FILING DATE: 13-AUG-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 214509/1996  
; FILING DATE: 14-AUG-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 199979/1997  
; FILING DATE: 25-JUL-1997

ATTORNEY/AGENT INFORMATION:  
NAME: HAILE, LISA A.  
REGISTRATION NUMBER: 38,347  
REFERENCE/DOCKET NUMBER: 07898/016001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619-678-5070  
TELEFAX: 619-678-5099  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 117 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-910-856-4

Query Match 4.9%; Score 6; DB 2; Length 117;  
Best Local Similarity 100.0%; Pred. No. 38;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 67 KKSTGK 72  
|||||  
Db 111 KKSTGK 116

RESULT 12  
US-08-957-130-15  
; Sequence 15, Application US/08957130  
; Patent No. 6290959  
; GENERAL INFORMATION:  
; APPLICANT: WU, Xue-Ru  
; APPLICANT: SUN, Tung-Tien  
; TITLE OF INVENTION: METHOD FOR SCREENING COMPOUNDS FOR  
; FILE OF INVENTION: INHIBITING BACTERIAL ATTACHMENT TO HOST CELL RECEPTORS  
; NUMBER OF SEQUENCES: 20  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BROWDY AND NEIMARK, P.L.L.C.  
; STREET: 419 Seventh Street N.W., Ste. 300  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20004  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/957,130  
; FILING DATE:  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BROWDY, Roger L.  
; REGISTRATION NUMBER: 25,618  
; REFERENCE/DOCKET NUMBER: SUN-8  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202) 628-5197  
; TELEFAX: (202) 737-3528  
; INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 260 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-957-130-15

Query Match 4.9%; Score 6; DB 4; Length 260;  
Best Local Similarity 100.0%; Pred. No. 80;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 35 SLKEPL 40

```
Db 195 SLKEPL 200
|||||
RESULT 13
US-08-446-345-34
; Sequence 34, Application US/08446345
; Patent No. 5831009
; GENERAL INFORMATION:
; APPLICANT: Ullrich, Axel
; APPLICANT: Moller, Niels P.H.
; APPLICANT: Moller, Karin B.
; TITLE OF INVENTION: NOVEL PROTEIN PHOSPHOTYROSINE
; TITLE OF INVENTION: PHOSPHATASES PTP-D1
; NUMBER OF SEQUENCES: 41
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: N.Y.
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/446,345
; FILING DATE: 22-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/234,440
; FILING DATE: 28-APR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A.
; REGISTRATION NUMBER: 30742
; REFERENCE/DOCKET NUMBER: 7683-054
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)790-9090
; TELEFAX: (212) 869-8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 34:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 272 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-08-446-345-34

Query Match 4.9%; Score 6; DB 2; Length 272;
Best Local Similarity 100.0%; Pred. No. 84;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 111 PLLVHC 116
Db 201 PLLVHC 206
|||||
RESULT 14
US-08-907-674-3
; Sequence 3, Application US/08907674
; Patent No. 5919685
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Shah, Purvi
; APPLICANT: Guegler, Karl J.
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: HUMAN AFLATOXIN B1 ALDEHYDE REDUCTASE
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/215,087
; FILING DATE:

Qy 85 PDHGTG 90
Db 111 PDHGTG 116
|||||
RESULT 15
US-09-215-087-3
; Sequence 3, Application US/09215087
; Patent No. 5981244
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Shah, Purvi
; APPLICANT: Guegler, Karl J.
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: HUMAN AFLATOXIN B1 ALDEHYDE REDUCTASE
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/215,087
; FILING DATE:

; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/907,674
; FILING DATE: 514
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0362 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 327 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 433611
; US-08-907-674-3
```

```

: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/907,674
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Billings, Lucy J.
: REGISTRATION NUMBER: 36,749
: REFERENCE/DOCKET NUMBER: PF-0362 US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 415-855-0555
: TELEFAX: 415-845-4166
: TELEX:
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 327 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: IMMEDIATE SOURCE:
: LIBRARY: GenBank
: CLONE: 433611
: US-09-215-087-3

```

```

Query Match 4.9%; Score 6; DB 2; Length 327;
Best Local Similarity 100.0%; Pred. No. 1e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 85 PDHCTP 90
Db 111 PDHCTP 116

```

Search completed: March 30, 2002, 08:27:44  
Job time: 311 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model  
Run on: March 30, 2002, 08:17:58 ; Search time 19.18 seconds  
(without alignments)  
484.531 Million cell updates/sec

Title: US-09-095-478A-8  
Perfect score: 122  
Sequence: 1 DFWGMWNNCNVAMITRE.....VRKSHITGPLLHCHTAGVGR 122

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 219241 seqs, 76174552 residues  
Word size : 0  
Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : PIR\_68:.\*  
1: pir1:.\*  
2: pir2:.\*  
3: pir3:.\*  
4: pir4:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	10	8.2	398	2 T08716	protein-tyrosine-p
2	7	5.7	310	2 T22013	hypothetical prote
3	7	5.7	377	1 A48711	protein-tyrosine-p
4	7	5.7	397	2 T21154	hypothetical prote
5	7	5.7	488	2 G71568	probable (pet112)
6	7	5.7	488	2 B81722	glutamyl-tRNA(Gln)
7	7	5.7	1286	2 T33476	hypothetical prote
8	7	5.7	1301	1 A41622	protein-tyrosine-p
9	6	4.9	41	2 T12917	hypothetical prote
10	6	4.9	49	2 H81991	hypothetical prote
11	6	4.9	83	2 S37764	hypothetical prote
12	6	4.9	91	2 C33172	C-ORF-F protein -
13	6	4.9	94	2 D70245	hypothetical prote
14	6	4.9	98	2 G84296	hypothetical prote
15	6	4.9	116	2 G71168	hypothetical prote
16	6	4.9	118	2 S51973	hypothetical prote
17	6	4.9	124	2 T50095	MHC class II beta
18	6	4.9	129	2 S36999	ribosomal protein
19	6	4.9	144	2 B59304	methyl viologen-re
20	6	4.9	152	1 M6URIC	Spec1 protein - se
21	6	4.9	164	2 T00891	hypothetical prote
22	6	4.9	167	2 C64456	hypothetical prote
23	6	4.9	170	2 S76087	hypothetical prote
24	6	4.9	174	2 S73301	hypothetical prote
25	6	4.9	175	2 T14679	repressor of phase
26	6	4.9	185	2 D71854	fkbp-type peptidyl
27	6	4.9	185	2 C64660	peptidyl-prolyl ci
28	6	4.9	187	2 T14935	hypothetical prote
29	6	4.9	189	2 S44639	hypothetical prote

ALIGNMENTS

RESULT : 1

T08716  
protein-tyrosine-phosphatase homolog DKF2p566K0524.1 - human (fragment)  
C:Species: Homo sapiens (man)  
C:Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 13-Aug-1999  
C:Accession: T08716  
R:Ansorge, W.; Wirkner, U.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.  
submitted to the Protein Sequence Database, May 1999  
A:Reference number: Z16472  
A:Accession: T08716  
A:Molecule type: mRNA  
A:Residues: 1-398 <ANS>  
A:Cross-references: EMBL:AL050040  
A:Experimental source: fetal kidney; clone DKF2p566K0524  
C:Genetics:  
C:Note: DKF2p566K0524.1  
C:Superfamily: protein-tyrosine-phosphatase homolog  
F:161-379/Domain: protein-tyrosine-phosphatase homology <PTP>

Query Match 8.2%; Score 10; DB 2; Length 398;  
Best Local Similarity 100.0%; Pred. No. 0.0059;  
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 83 KWPDHGTPAS 92  
DB 298 KWPDHGTPAS 307

RESULT : 2

T22013  
hypothetical protein F40D4.5 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 20-Jun-2000  
C:Accession: T22013  
R:Matthews, L.  
submitted to the EMBL Data Library, November 1996  
A:Reference number: Z19502  
A:Accession: T22013  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-310 <WILL>  
A:Cross-references: EMBL:Z81536; PIDN:CA804364.1; GSPDB:GN00023; CESP:F40D4.5  
A:Experimental source: clone F40D4  
C:Genetics:  
A:Gene: CESP:F40D4.5  
A:Map position: 5  
A:Introns: 45/1; 114/2; 194/1  
C:Superfamily: Caenorhabditis elegans hypothetical protein C13D9.6

Query Match 5.7%; Score 7; DB 2; Length 310;

Best Local Similarity 100.0%; Pred. No. 9.5;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 108 ITGPLL 114  
|||||  
Db 234 ITGPLL 240

## RESULT 3

A48711  
A:Title: protein-tyrosine-phosphatase (BC 3.1.3.48), nonreceptor type 2 - slime mold (Dictyostelium)  
N:Alternate names: protein-tyrosine-phosphatase DdPTPa; PTP2  
C:Species: Dictyostelium discoideum  
C:Date: 02-Jun-1995 #sequence\_revision 08-Mar-1996 #text\_change 11-Jun-1999  
C:Accession: A48711; A53774  
R:Ramalingam, R.; Shaw, D.R.; Ennis, H.L.  
J. Biol. Chem. 268, 22680-22685, 1993  
A:Title: Cloning and functional expression of a Dictyostelium discoideum protein tyrosine  
A:Reference number: A48711; MUID:94043028  
A:Accession: A48711  
A:Molecule type: mRNA  
A:Residues: 1-377 <RAM>  
A:Cross-references: GB:L15420; NID:g290036; PIDN:AAA33242.1; PID:g290037  
R:Howard, P.K.; Gamber, M.; Hunter, T.; Firtel, R.A.  
Mol. Cell. Biol. 14, 5154-5164, 1994  
A:Title: Regulation by protein-tyrosine phosphatase PTP2 is distinct from that by PTP1  
A:Reference number: A53774; MUID:94309635  
A:Accession: A53774  
A:Status: nucleic acid sequence not shown  
A:Molecule type: mRNA; DNA  
A:Residues: 1-377 <HOW>  
A:Cross-references: GB:L15420; NID:g290036; PIDN:AAA33242.1; PID:g290037  
C:Note: the translation of the nucleotide sequence is not complete in this paper  
C:Superfamily: Dictyostelium protein-tyrosine-phosphatase, nonreceptor type 2; protein-  
F:114-336/Domain: protein-tyrosine-phosphatase homology #status atypical <PTP>  
F:281/Active site: Cys (phosphocysteine intermediate) #status predicted  
F:287/Binding site: substrate phosphate (Arg) #status predicted

Query Match 5.7%; Score 7; DB 1; Length 377;  
Best Local Similarity 100.0%; Pred. No. 11;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 WPDHGTP 90  
|||||  
Db 249 WPDHGTP 255

## RESULT 4

T21154  
A:Title: hypothetical protein F20E11.2 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 21-Jul-2000  
C:Accession: T21154  
R:Ainscough, R.  
submitted to the EMBL Data Library, November 1996  
A:Reference number: Z19383  
A:Accession: T21154  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-397 <WIL>  
A:Cross-references: EMBL:Z19383; PIDN:CAB04140.1; GSPDB:GN00023; CRSP:F20E11.2  
A:Experimental source: clone F20E11  
C:Genetics:  
A:Gene: CRSP:F20E11.2  
A:Map position: 5  
A:Introns: 45/1; 114/2; 194/1; 281/1; 316/3  
C:Superfamily: Caenorhabditis elegans hypothetical protein C13D9.6

Query Match 5.7%; Score 7; DB 2; Length 397;  
Best Local Similarity 100.0%; Pred. No. 12;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 108 ITGPLL 114  
|||||  
Db 234 ITGPLL 240

## RESULT 5

G71568  
A:Title: probable (pet112) glu-tRNA gln amidotransferase (b chain) - Chlamydia trachomatis (se  
C:Species: Chlamydia trachomatis  
C:Date: 13-Sep-1998 #sequence\_revision 13-Sep-1998 #text\_change 08-Oct-1999  
C:Accession: G71568  
R:Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitche  
Science 282, 754-759, 1998  
A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia t  
A:Reference number: A71570; MUID:99000809  
A:Accession: G71568  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-488 <ARN>  
A:Cross-references: GB:AE001275; GB:AE001273; NID:g3328388; PIDN:AAC67594.1; PID:g332  
A:Experimental source: serotype D, strain UW-3/Cx  
C:Genetics:  
A:Gene: gatB  
C:Superfamily: PET112 protein

Query Match 5.7%; Score 7; DB 2; Length 488;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 51 NFQITQY 57  
|||||  
Db 94 NFQITQY 100

## RESULT 6

B81722  
A:Title: glutamyl-tRNA(Gln) amidotransferase chain B TC0272 [Imported] - Chlamydia muridarum (C  
C:Species: Chlamydia muridarum, Chlamydia trachomatis MoPn  
C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 19-May-2000  
C:Accession: B81722  
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hicke  
C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzbe  
Nucleic Acids Res. 28, 1397-1406, 2000  
A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39  
A:Reference number: A81500; MUID:20150255  
A:Accession: B81722  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-488 <TET>  
A:Cross-references: GB:AE002294; GB:AE002160; NID:g7190305; PIDN:AAF39140.1; PID:g719  
A:Experimental source: strain Nigg (MoPn)  
C:Genetics:  
A:Gene: TC0272  
C:Superfamily: PET112 protein

Query Match 5.7%; Score 7; DB 2; Length 488;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 51 NFQITQY 57  
|||||  
Db 94 NFQITQY 100

## RESULT 7

T33476  
A:Title: hypothetical protein T27C10.6 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999

C:Accession: T33476  
 R:Zhu, H.J.; Graves, T.; Hawkins, M.  
 submitted to the EMBL Data Library, October 1998  
 A:Description: The sequence of *C. elegans* cosmid T27C10.  
 A:Reference number: 221354  
 A:Accession: T33476  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-1286 <ZHU>  
 A:Cross-references: EMBL:AF098504; PIDN:AAC67414.1; GSPDB:GN00019; CESP:T27C10.6  
 A:Experimental source: strain Bristol N2; clone T27C10  
 C:Genetics:  
 A:Gene: CESP:T27C10.6  
 A:Map position: 1  
 A:Introns: 26/3; 79/1; 145/1; 182/2; 203/2; 269/1; 341/2; 397/3; 437/2; 719/1; 835/3

Query Match 5.7%; Score 7; DB 2; Length 1286;  
 Best Local Similarity 100.0%; Pred. No. 33;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 35 SLKEPLE 41  
 |||||  
 Db 998 SLKEPLE 1004

RESULT 8  
 A1622  
 protein-tyrosine-phosphatase (EC 3.1.3.48), receptor type 99A precursor - fruit fly (*Drosophila melanogaster*)  
 N:Alternate names: phosphotyrosine phosphatase 99A  
 C:Species: *Drosophila melanogaster*  
 C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
 C:Accession: A1622; A1214; B41214; B41215  
 R:Harikaran, I.K.; Chuang, P.T.; Rubin, G.M.  
 Proc. Natl. Acad. Sci. U.S.A. 88, 11266-11270, 1991  
 A:Title: Cloning and characterization of a receptor-class phosphotyrosine phosphatase gene  
 A:Reference number: A41622; MUID:92107930  
 A:Accession: A41622  
 A:Molecule type: mRNA  
 A:Residues: 1-1301 <HAR>  
 A:Cross-references: GB:M81795; NID:g157293; PIDN:AAA28483.1; PID:g157294  
 R:Yang, X.; Seow, K.T.; Bahri, S.M.; Oon, S.H.; Chia, W.  
 Cell 67, 661-673, 1991  
 A:Title: Two *Drosophila* receptor-like tyrosine phosphatase genes are expressed in a subset of neurons  
 A:Reference number: A41214; MUID:92034988  
 A:Accession: A41214  
 A:Molecule type: mRNA  
 A:Residues: 1-585, 'R', 587-1049, 1120-1204, 'H', 1206-1301 <YAN>  
 A:Cross-references: GB:M80464; NID:g157299; PIDN:AAA28486.1; PID:g157300  
 A:Accession: B41214  
 A:Molecule type: mRNA  
 A:Residues: 1-585, 'R', 587-1049, 1290-1301 <YA2>  
 A:Cross-references: GB:M80464  
 R:Tian, S.S.; Tsoulfas, P.; Zinn, K.  
 Cell 67, 675-685, 1991  
 A:Title: Three receptor-linked protein-tyrosine phosphatases are selectively expressed in different subsets of neurons  
 A:Reference number: A41215; MUID:92034989  
 A:Accession: B41215  
 A:Molecule type: mRNA  
 A:Residues: 1-585, 'R', 587-1049, 1120-1184, 'S', 1186-1301 <TIA>  
 A:Cross-references: GB:M80539  
 C:Genetics:  
 A:Gene: FlyBase:Ptp99A  
 A:Cross-references: FlyBase:FBgn0004369  
 C:Superfamily: protein-tyrosine-phosphatase, receptor type 99A; fibronectin type III repeat domain  
 C:Keywords: alternative splicing; phosphoprotein; phosphoric monoester hydrolase; transmembrane protein  
 F:1-29/Domain: signal sequence #status predicted <SIG>  
 F:30-1301/product: protein-tyrosine-phosphatase, receptor type 99A #status predicted <MA>  
 F:30-1049, 1120-1301/product: protein-tyrosine-phosphatase, receptor type 99A, medium splice variant  
 F:30-1049, 1290-1301/product: protein-tyrosine-phosphatase, receptor type 99A, short splice variant  
 F:393-416/Domain: transmembrane #status predicted <TM>  
 F:502-730/Domain: protein-tyrosine-phosphatase homology <PTP>  
 F:799-1005/Domain: protein-tyrosine-phosphatase homology <PTP2>

F:1059-1091/Region: glutamine-rich  
 F:682/Active site: Cys (phosphocysteine intermediate) #status predicted  
 F:688/Binding site: substrate phosphate (Arg) #status predicted

Query Match 5.7%; Score 7; DB 1; Length 1301;  
 Best Local Similarity 100.0%; Pred. No. 33;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 84 WPDHGTP 90  
 |||||  
 Db 648 WPDHGTP 654

RESULT 9  
 T12917  
 hypothetical protein yosF - *Bacillus subtilis* phage SPBc2  
 C:Species: *Bacillus subtilis* phage SPBc2  
 C:Date: 13-Aug-1999 #sequence\_revision 13-Aug-1999 #text\_change 15-Oct-1999  
 C:Accession: T12917; H69925  
 R:Lazarévic, V.; Duesterhoeft, A.; Soldo, B.; Hilbert, H.; Maue, C.; Karamata, D.  
 submitted to the EMBL Data Library, August 1997  
 A:Description: The complete nucleotide sequence of the *Bacillus subtilis* SPBc2 protein  
 A:Reference number: 217583  
 A:Accession: T12917  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-41 <LA2>  
 A:Cross-references: EMBL:AF020713; NID:g3025478; PID:g3025631; PIDN:AAC13126.1  
 R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber  
 C.: Brop, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;  
 C.: Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,  
 Nature 390, 249-256, 1997  
 A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal  
 lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hulio, M  
 Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino  
 A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau  
 Y. M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portet  
 Rieger, M.; Rivoita, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanl  
 A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se  
 akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiya  
 T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida  
 A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.  
 A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*  
 A:Reference number: A69580; MUID:98044033  
 A:Accession: H69925  
 A:Status: nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-41 <KUN>  
 A:Cross-references: GB:299114; GB:AL009126; NID:g2634230; PIDN:CAB13906.1; PID:ell1854  
 A:Experimental source: strain 168  
 C:Genetics:  
 A:Gene: yosF

Query Match 4.9%; Score 6; DB 2; Length 41;  
 Best Local Similarity 100.0%; Pred. No. 21;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 67 KKSTCK 72  
 |||||  
 Db 63 KKSTCK 18

RESULT 10  
 H8191  
 hypothetical protein NMA0012 [imported] - *Neisseria meningitidis* (strain 22491 serogr  
 C:Species: *Neisseria meningitidis*  
 C:Date: 05-May-2000 #sequence\_revision 05-May-2000 #text\_change 02-Feb-2001  
 C:Accession: H8191  
 R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Mo  
 Holroyd, S.; Jagsels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandre  
 Nature 404, 502-506, 2000

A:Title: Complete DNA sequence of a serogroup A strain of *Neisseria meningitidis* Z2491.  
A:Reference number: A81775; MUID:20222536  
A:Accession: H81991  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-49 <PAR>  
A:Cross-references: GB:AL157959; NID:g7378778; PIDN:CAB83332.1; PID:g737879  
A:Experimental source: serogroup A, strain Z2491  
C:Genetics:  
A:Gene: NMA0012

Query Match 4.9%; Score 6; DB 2; Length 49;  
Best Local Similarity 100.0%; Pred. No. 24;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 117 TAGVGR 122  
|||||  
Db 15 TAGVGR 20

RESULT 11  
S37764  
hypothetical protein - fruit fly (*Drosophila miranda*) transposon TRIM  
C:Species: *Drosophila miranda*  
C:Date: 19-May-1994 #sequence\_revision 01-Dec-1995 #text\_change 21-Jul-2000  
C:Accession: S37764  
R:Steinmann, M.; Steinmann, S.  
A:Title: Preferential X chromosomal location of TRIM, a novel transposable element of *Drosophila*  
A:Reference number: S37763; MUID:92164374  
A:Accession: S37764  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-83 <STE>  
A:Cross-references: EMBL:X59239; NID:g8718; PIDN:CAA41924.1; PID:g1335690  
C:Genetics:  
A:Gene: FlyBase:Dmir/TRIM  
A:Cross-references: FlyBase:FBgn0004642  
A:Mobile element: transposon TRIM

Query Match 4.9%; Score 6; DB 2; Length 83;  
Best Local Similarity 100.0%; Pred. No. 38;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 19 REIEGG 24  
|||||  
Db 36 REIEGG 41

RESULT 12  
C33172  
C-ORF-F protein - vaccinia virus (strain Copenhagen)  
N:Alternate names: B-ORF-G protein  
C:Species: vaccinia virus  
A:Note: host *Homo sapiens* (man)  
C:Date: 19-Apr-1991 #sequence\_revision 19-Apr-1991 #text\_change 05-Jan-1996  
C:Accession: C33172; H42529  
R:Johnson, G.P.  
A:Reference number: A33172  
A:Accession: C33172  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-91 <JOH1>  
A:Accession: H42529  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-91 <JOH2>  
C:Genetics:  
A:Note: the coding region for this protein is repeated in the viral genome

Query Match 4.9%; Score 6; DB 2; Length 91;  
Best Local Similarity 100.0%; Pred. No. 41;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 98 KYRVYV 103  
|||||  
Db 31 KYRVYV 36

RESULT 13  
D70245  
hypothetical protein BBJ02 - Lyme disease spirochete plasmid J/lp38  
C:Species: *Borrelia burgdorferi* (Lyme disease spirochete)  
C:Date: 13-Feb-1998 #sequence\_revision 13-Feb-1998 #text\_change 08-Oct-1999  
C:Accession: D70245  
R:Fraser, C.M.; Casjens, S.; Huang, W.M.; Sutton, G.G.; Clayton, R.; Lathigra, R.; Wh  
son, D.; Peterson, J.; Kerlavage, A.R.; Quackenbush, J.; Salzberg, S.; Hanson, M.; Vu  
; Bowman, C.; Garland, S.; Fujii, C.; Cotton, M.D.; Horst, K.; Roberts, K.; Hatch, B.  
Nature 390, 580-586, 1997  
A:Authors: Smith, H.O.; Venter, J.C.  
A:Title: Genomic sequence of a Lyme disease spirochaete, *Borrelia burgdorferi*.  
A:Reference number: A70100; MUID:98065943  
A:Accession: D70245  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-94 <KLE>  
A:Cross-references: GB:AE000787; NID:g2690175; PIDN:AAC66115.1; PID:g3690207; TIGR:BB  
A:Experimental source: strain B31  
C:Genetics:  
A:Genome: plasmid

Query Match 4.9%; Score 6; DB 2; Length 94;  
Best Local Similarity 100.0%; Pred. No. 42;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 64 QIVKKS 69  
|||||  
Db 73 QIVKKS 78

RESULT 14  
G84296  
hypothetical protein Vng1425h [imported] - *Halobacterium* sp. NRC-1  
C:Species: *Halobacterium* sp. NRC-1  
C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
C:Accession: G84296  
R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky  
; Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Ja  
Jung, K.H.; Alam, M.; Freitas, T.  
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000  
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.;  
A:Title: Genome sequence of *Halobacterium* species NRC-1.  
A:Reference number: A84160; MUID:20504483  
A:Accession: G84296  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-98 <STO>  
A:Cross-references: GB:AE004437; NID:g10580927; PIDN:AAG19739.1; GSPDB:GN00138  
C:Genetics:  
A:Gene: VNG1425H

Query Match 4.9%; Score 6; DB 2; Length 98;  
Best Local Similarity 100.0%; Pred. No. 44;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 90 PASVDF 95  
|||||  
Db 49 PASVDF 54



RESULT 15

G71168  
 Hypothetical protein PH0547 - Pyrococcus horikoshii  
 C:Species: Pyrococcus horikoshii  
 C:Date: 14-Aug-1998 #sequence\_revision 14-Aug-1998 #text\_change 20-Jun-2000  
 C:Accession: G71168  
 R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekin  
 M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi  
 DNA Res. 5, 55-76, 1998  
 A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic a  
 A:Reference number: A71000; MUID:98344137  
 A:Accession: G71168  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-116 <KAW>  
 A:Cross-references: GB:AP000002; NID:g3236129; PIDN:BAA29636.1; PID:g3256953  
 A:Experimental source: strain OT3  
 A:Note: this accession replaces an interim accession for a sequence replaced by GenBank  
 C:Genetics:  
 A:Gene: PH0547  
 C:Superfamily: Pyrococcus horikoshii hypothetical protein PH0547

Query Match 4.9%; Score 6; DB 2; Length 116;  
 Best Local Similarity 100.0%; Pred. No. 51;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 14 IAMITR 19  
 |||||  
 Db 39 IAMITR 44

Search completed: March 30, 2002, 08:27:17  
 Job time: 559 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 30, 2002, 08:27:19 ; Search time 10.3 seconds  
(without alignments)  
434.282 Million cell updates/sec

Title: US-09-095-478A-8  
Perfect score: 122  
Sequence: 1 DFWMWNNCNCVIAITRE.....VRKSHITGPLLHVHCTAGVGR 122

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0  
Searched: 100059 seqs, 36664827 residues

Word size : 0

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : SwissProt\_39:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Length	DB ID	Description
1	7	5.7	112	1 PT07_STYPL	P28199 styela plic
2	7	5.7	377	1 PTP2_DICDI	P34138 dictyosteli
3	7	5.7	488	1 GATB_CHLMU	O9p136 chlamydia m
4	7	5.7	488	1 GATB_CHLTR	O84007 chlamydia t
5	7	5.7	1301	1 PTP9_DROME	P35832 drosophila
6	7	5.7	2472	1 NCR2_MOUSE	O9wu42 mus musculu
7	6	4.9	91	1 YVBG_VACCC	P20547 vaccinia vi
8	6	4.9	114	1 PT27_STYPL	P28219 styela plic
9	6	4.9	117	1 PT25_STYPL	P28217 styela plic
10	6	4.9	118	1 YAE6_YEAST	P39724 saccharomyc
11	6	4.9	129	1 RL31_SULAC	P38618 sulfolobus
12	6	4.9	144	1 CSF2_FEICA	O62757 felis silve
13	6	4.9	152	1 SPIA_STRPU	P04109 strongyloce
14	6	4.9	174	1 YC21_PORPU	P51380 porphyra pu
15	6	4.9	185	1 SLYD_HELPJ	O92k89 helicobacte
16	6	4.9	185	1 SLYD_HELPY	O25748 helicobacte
17	6	4.9	189	1 YPT2_CAEEL	P41880 caenorhabdi
18	6	4.9	219	1 Y284_TREPA	O83308 treponema p
19	6	4.9	243	1 CY1_EUGGR	P20114 euglena gra
20	6	4.9	251	1 HIS6_SULSO	O33774 sulfolobus
21	6	4.9	259	1 UPKB_BOVIN	P38573 bos taurus
22	6	4.9	259	1 UPKB_MUSVI	P30413 mustela vis
23	6	4.9	279	1 ATPC_MYCGE	P47640 mycoplasma
24	6	4.9	306	1 YEDA_ECOLI	P09185 escherichia
25	6	4.9	309	1 Y209_MYCPN	P75485 mycoplasma
26	6	4.9	323	1 CRO_RANCA	P17264 rana catesb
27	6	4.9	323	1 CRO_RANTE	P02532 rana tempor
28	6	4.9	327	1 AR71_RAT	P38918 rattus norv
29	6	4.9	330	1 AR72_HUMAN	O43488 homo sapien
30	6	4.9	339	1 REP_BACAM	P13963 bacillus am
31	6	4.9	371	1 Y028_ARCFU	O30207 archaeoglob
32	6	4.9	379	1 METX_MYCTU	O53391 mycobacteri
33	6	4.9	382	1 METX_MYCLE	O32874 mycobacteri

34	6	4.9	391	1 TAL_SYNY3	P72797 synecocyst
35	6	4.9	395	1 RRP_NDVA	P16073 newcastle d
36	6	4.9	409	1 YM65_MYCTU	O50697 mycobacteri
37	6	4.9	421	1 RF1_METJA	O58239 methanococc
38	6	4.9	429	1 MYCS_RAT	P23999 rattus norv
39	6	4.9	440	1 RHGA_ASPAC	O00001 aspergillus
40	6	4.9	445	1 GSPB_BACST	P13376 bacillus st
41	6	4.9	469	1 CS66_WHEAT	P46526 triticum ae
42	6	4.9	471	1 ATPB_MYCGA	P33253 mycoplasma
43	6	4.9	476	1 MM10_MOUSE	O55123 mus musculu
44	6	4.9	477	1 MM03_MOUSE	P28862 mus musculu
45	6	4.9	487	1 DHAB_RHINE	P54222 rhizobium m

ALIGNMENTS

RESULT 1  
PT07\_STYPL STANDARD; PRT; 112 AA.  
AC P28199;  
DT 01-DEC-1992 (Rel. 24, Created)  
DT 30-MAY-2000 (Rel. 39, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE PROTEIN-TYROSINE PHOSPHATASE 7 (EC 3.1.3.48) (FRAGMENT).  
GN Sty 7.  
OS Styela plicata (Sea squirt).  
OC Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea;  
OC Stolidobranchia; Styelidae; Styela.  
OX NCBI\_TaxID=7726;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91139172; PubMed=1704870;  
FA Matthews R.J., Flores E., Thomas M.L.;  
RT "Protein tyrosine phosphatase domains from the protochordate Styela plicata".  
RL Immunogenetics 33:33-41(1991).  
RN [2]  
RP REVISIONS TO C-TERMINUS.  
FA Matthews R.J., Flores E., Thomas M.L.;  
PL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.  
CC -1- CATALYTIC ACTIVITY: PROTEIN TYROSINE PHOSPHATE + H(2)O -> PROTEIN TYROSINE + ORTHOPHOSPHATE.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; M37992; AAA29825.1; -  
DR HSSP; P28827; 1RPM.  
DR InterPro; IPR000340; DS\_phosphatase.  
DR InterPro; IPR000387; TYR\_phosphatase.  
DR InterPro; IPR000242; Tyr\_prot\_phptase.  
DR Pfam; PF00102; Y\_phosphatase; 1.  
DR PRINTS; PR00700; PRTYPHPTASE.  
DR SMART; SM00012; PTPC\_DSPC; 1.  
DR PROSITE; PS00383; TYR\_PHOSPHATASE\_1; PARTIAL.  
DR PROSITE; PS50056; TYR\_PHOSPHATASE\_2; PARTIAL.  
DR PROSITE; PS50055; TYR\_PHOSPHATASE\_PTP; 1.  
KW Hydrolase.  
FT NON\_TER  
FT NON\_TER 112  
FT NON\_TER 112  
SQ SEQUENCE 112 AA; 13299 MW; 6F19D514C685D9FC CRC64;

Query Match 5.7%; Score 7; DB 1; Length 112;  
Best local Similarity 100.0%; Pred. No. 1.6;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 84 WPDHGT 90
DB 80 WPDHGT 86

RESULT 2
PTP2_DICDI STANDARD; PRT; 377 AA.
AC P34138;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE PROTEIN-TYROSINE PHOSPHATASE 2 (EC 3.1.3.48) (PROTEIN-TYROSINE-
DE PHOSPHATE PHOSPHOHYDROLASE 2) (PTPA).
GN PTPB OR PTP2.
OS Dictyostelium discoideum (Slime mold).
OS Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
ON NCBI_TaxID=4689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AX4;
RX MEDLINE=94043028; PubMed=8226777;
RA Ramalingam R., Shaw D.R., Ennis H.;
RT "Cloning and functional expression of a Dictyostelium discoideum
RT protein tyrosine phosphatase.";
RL J. Biol. Chem. 268:22680-22685(1993).
CC -!- CATALYTIC ACTIVITY: PROTEIN TYROSINE PHOSPHATE + H(2)O =
CC PROTEIN TYROSINE + ORTHOPHOSPHATE.
CC -!- SIMILARITY: BELONGS TO THE NON-RECEPTOR CLASS OF THE PROTEIN-
CC TYROSINE PHOSPHATASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L15420; AAA33242.1; -
DR HSP; P18031; 2HNP.
DR DictyDB; D05054; ptpb.
DR InterPro; IPR000387; TYR_phosphatase.
DR Pfam; PF00102; Y_phosphatase.1.
DR PRINTS; PR00700; PTPYPHPTASE.
DR SMART; SM00194; PTPC; 1.
DR PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
DR PROSITE; PS50056; TYR_PHOSPHATASE_2; 1.
DR PROSITE; PS50055; TYR_PHOSPHATASE_PTP; 1.
KW Hydrolase.
FT DOMAIN 78 86 POLY-ASP.
FT DOMAIN 95 103 POLY-ASN.
FT ACT_SITE 281 281 BY SIMILARITY.
SQ SEQUENCE 377 AA; 43488 MW; 5B08F6EB54829FD9 CRC64;

Query Match 5.7%; Score 7; DB 1; Length 377;
Best Local Similarity 100.0%; Pred. No. 4.6;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 WPDHGT 90
DB 249 WPDHGT 255

RESULT 3
GATB_CHLMU STANDARD; PRT; 488 AA.
AC Q9PL36;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)

QY 84 WPDHGT 90
DB 249 WPDHGT 255

RESULT 4
GATB_CHLTR STANDARD; PRT; 488 AA.
AC Q84007;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE GLUTAMYL-TRNA(GLN) AMIDOTRANSFERASE SUBUNIT B (EC 6.3.5.-) (GLU-ADT
DE SUBUNIT B).
GN GATB OR CT004.
OS Chlamydia trachomatis.
OS Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
ON NCBI_TaxID=813;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=D/UW-3/CX;
RX MEDLINE=99000809; PubMed=9784136;
RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,

DE GLUTAMYL-TRNA(GLN) AMIDOTRANSFERASE SUBUNIT B (EC 6.3.5.-) (GLU-ADT
DE SUBUNIT B).
GN GATB OR TC0272.
OS Chlamydia muridarum.
OS Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
ON NCBI_TaxID=83560;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MOPN / NIGG;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Uterback T., Berry K., Bass S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis Mopn and Chlamydia
RT pneumoniae AR39.";
RL Nucleic Acids Res. 28:1397-1406(2000).
CC -!- FUNCTION: FURNISHES A MEANS FOR FORMATION OF CORRECTLY CHARGED
CC GLN-TRNA(GLN) THROUGH THE TRANSMIDATION OF MISACYLATED GLU-
CC TRNA(GLN) IN ORGANISMS WHICH LACK GLUTAMINYL-TRNA SYNTHETASE. THE
CC REACTION TAKES PLACE IN THE PRESENCE OF GLUTAMINE AND ATP THROUGH
CC AN ACTIVATED GAMMA-PHOSPHO-GLU-TRNA(GLN) (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: ATP + L-GLUTAMYL-TRNA(GLN) + L-GLUTAMINE = ADP
CC + PHOSPHATE + L-GLUTAMINYL-TRNA(GLN) + L-GLUTAMATE.
CC -!- SUBUNIT: HETEROTRIMER OF A, B AND C SUBUNITS (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE GATB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE002294; AAF39140.1; -
DR TIGR; TC0272; -
DR InterPro; IPR001773; Gln_amidotransf_B.
DR Pfam; PF01162; PET112; 1.
DR PROSITE; PS01234; GATB; 1.
KW Protein biosynthesis; Ligase; Complete proteome.
SQ SEQUENCE 488 AA; 54759 MW; B1966520739D041E CRC64;

Query Match 5.7%; Score 7; DB 1; Length 488;
Best Local Similarity 100.0%; Pred. No. 5.8;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 51 NFOITQY 57
DB 94 NFOITQY 100

```

RA Davis R.W.;

RT "Genome sequence of an obligate intracellular pathogen of humans:

RL Chlamydia trachomatis.;"

CC Science 282:754-759(1998).

CC -1- FUNCTION: FURNISHES A MEANS FOR FORMATION OF CORRECTLY CHARGED

CC GLN-TRNA(GLN) THROUGH THE TRANSMINATION OF MISACYLATED GLU-

CC TRNA(GLN) IN ORGANISMS WHICH LACK GLUTAMINYL-TRNA SYNTHETASE. THE

CC REACTION TAKES PLACE IN THE PRESENCE OF GLUTAMINE AND ATP THROUGH

CC AN ACTIVATED GAMMA-PHOSPHO-GLU-TRNA(GLN) (BY SIMILARITY).

CC -1- CATALYTIC ACTIVITY: ATP + L-GLUTAMYL-TRNA(GLN) + L-GLUTAMINE = ADP

CC + PHOSPHATE + L-GLUTAMINYL-TRNA(GLN) + L-GLUTAMATE.

CC -1- SUBUNIT: HETEROTRIMER OF A, B AND C SUBUNITS (BY SIMILARITY).

CC -1- SIMILARITY: BELONGS TO THE GATB FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC EMBL; AE001275; AAC67594.1; -

DR InterPro; IPR001773; Gln amidotransf\_B.

DR Pfam; PF01162; PET112; 1

DR PROSITE; PS01234; GATB; 1.

KW Protein biosynthesis; ligase; Complete proteome.

SO SEQUENCE 488 AA; 55018 MW; A6F9D586042DA112 CRC64;

Query Match 5.7%; Score 7; DB 1; Length 488;

Best Local Similarity 100.0%; Pred. No. 5.8;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 51 NFIQITQY 57

Db 94 NFIQITQY 100

|||||

RESULT 5

PTP9\_DROME STANDARD; PRT; 1301 AA.

AC P35832;

DT 01-JUN-1994 (Rel. 29, Created)

DT 01-JUN-1994 (Rel. 29, Last sequence update)

DT 20-AUG-2001 (Rel. 40, Last annotation update)

DE PROTEIN-TYROSINE PHOSPHATASE 99A PRECURSOR (EC 3.1.3.48) (RECEPTOR-

DE LINKED PROTEIN-TYROSINE PHOSPHATASE 99A).

GN PTP99A.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephyroidea; Drosophilidae; Drosophila.

OX NCBI\_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Eye imaginal disk;

RX MEDLINE=92107930; PubMed=1662390;

RA Hariharan I.K.; Chuang P.-T.; Rubin G.M.;

RT "Cloning and characterization of a receptor-class phosphotyrosine

RT phosphatase gene expressed on central nervous system axons in

RT Drosophila melanogaster.;"

RL Proc. Natl. Acad. Sci. U.S.A. 88:11266-11270(1991).

RN [2]

RP SEQUENCE FROM N.A.

RC TISSUE=Embryo;

RX MEDLINE=92034989; PubMed=1657402;

RA Tian S.-S.; Tsoulfas P.; Zinn K.;

RT "Three receptor-linked protein-tyrosine phosphatases are selectively

RT expressed on central nervous system axons in the Drosophila embryo.;"

RL Cell 67:675-685(1991).

RN [3]

RP SEQUENCE FROM N.A.

TISSUE=Embryo;

MEDLINE=92034988; PubMed=1657401;

RA Yang X.; Seow K.T.; Bahri S.M.; Oon S.H.; Chia W.;

RT "Two Drosophila receptor-like tyrosine phosphatase genes are

RT expressed in a subset of developing axons and pioneer neurons in the

RT embryonic CNS.;"

RL Cell 67:661-673(1991).

CC -1- FUNCTION: MAY PLAY A KEY ROLE IN SIGNAL TRANSDUCTION AND GROWTH

CC OF AXONS.

CC -1- CATALYTIC ACTIVITY: PROTEIN TYROSINE PHOSPHATE + H(2)O =

CC PROTEIN TYROSINE + ORTHOPHOSPHATE.

CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.

CC -1- ALTERNATIVE PRODUCTS: TWO ISOMERS THAT DIFFER IN THEIR C-TERMINAL

CC TAILS ARE PRODUCED BY ALTERNATIVE SPLICING.

CC -1- TISSUE SPECIFICITY: SELECTIVELY EXPRESSED IN A SUBSET OF AXONS AND

CC PIONEER NEURONS IN THE EMBRYO.

CC -1- SIMILARITY: CONTAINS 3 FIBRONECTIN TYPE III-LIKE DOMAINS

CC -1- SIMILARITY: CONTAINS 2 PROTEIN-TYROSINE PHOSPHATASE DOMAINS.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC EMBL; M81795; AAA28483.1; -

DR EMBL; M80539; AAA28485.1; -

DR EMBL; M80464; AAA28486.1; -

DR PIR; A41622; A41622.

DR PIR; A41214; A41214.

DR PIR; B41214; B41214.

DR PIR; B41215; B41215.

DR HSSP; P18052; LYFO.

DR FlyBase; FBgn0004369; Ptp99A.

DR InterPro; IPR001777; FN\_III.

DR InterPro; IPR000387; Tyr\_phosphatase.

DR Pfam; PF00041; fn3; 2.

DR Pfam; PF00102; Y\_phosphatase; 2.

DR PRINTS; PR00014; ENTYPRII.

DR PRINTS; PR00700; PTPPHPTASE.

DR SMART; SM00060; FN3; 2.

DR SMART; SM00194; PTPC; 2.

DR PROSITE; PS00383; TYR\_PHOSPHATASE\_1; 1.

DR PROSITE; PS00056; TYR\_PHOSPHATASE\_2; 1.

DR PROSITE; PS00055; TYR\_PHOSPHATASE\_PTP; 2.

KW Transmembrane; Hydrolase; Duplication; Signal; Alternative splicing;

KW Repeat.

FT SIGNAL. 1 29

FT CHAIN 30 1301

FT DOMAIN 30 392

FT TRANSMEM 393 416

FT DOMAIN 417 1301

FT DOMAIN 64 168

FT DOMAIN 169 268

FT DOMAIN 269 368

FT DOMAIN 497 747

FT DOMAIN 748 975

FT ACT\_SITE 682 682

FT DOMAIN 1076 1091

FT CARBOHYD 33 33

FT CARBOHYD 176 176

FT CARBOHYD 212 212

FT CARBOHYD 278 278

FT CARBOHYD 322 322

FT CARBOHYD 336 336

FT CARBOHYD 1050 1119

FT VARSPIC 586 586

FT CONFLICT 1205 1205

FT SEQUENCE 1301 AA; 145336 MW; 8241E3E19A4CA5BD CRC64;

FT N-LINKED (GLCNAC. . .) (POTENTIAL).

FT N-LINKED (GLCNAC. . .) (POTENTIAL).

FT N-LINKED (GLCNAC. . .) (POTENTIAL).

FT N-LINKED (GLCNAC. . .) (POTENTIAL).

FT N-LINKED (GLCNAC. . .) (POTENTIAL).

FT N-LINKED (GLCNAC. . .) (POTENTIAL).

FT MISSING (IN SHORT ISOFORM).

FT P -> R (IN REF. 2 AND 3).

FT N -> H (IN REF. 3).

```
Query Match 5.7%; Score 7; DB 1; Length 1301;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 84 WPDHGT 90
DB 648 WPDHGT 654

RESULT 6
ID NCR2_MOUSE STANDARD; PRT; 2472 AA.
AC Q9WU42; Q9WU43; Q9WU41;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE NUCLEAR RECEPTOR CO-REPRESSOR 2 (N-COR2) (SILENCING MEDIATOR OF
DE RETINOIC ACID AND THYROID HORMONE RECEPTOR) (SMRT) (THYROID-,
DE RETINOIC ACID-RECEPTOR-ASSOCIATED CO-REPRESSOR) (T3 RECEPTOR-
DE ASSOCIATING FACTOR) (TRAC).
GN NCOR2 OR SMRT.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM ALPHA AND ISOFORM BETA).
RC TISSUE=Spleen, and Brain;
RA MEDLINE=99178941; PubMed=10077563;
RX Ordentlich P., Downes M., Xie W., Genin A., Spinner N.B., Evans R.M.;
RT "Unique forms of human and mouse nuclear receptor corepressor SMRT.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:2639-2644(1999).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM ALPHA).
RC TISSUE=Embryo;
RX MEDLINE=9919215; PubMed=10097068;
RA Park E.J., Schroen D.J., Yang M., Li H., Li L., Chen J.D.;
RT "SMRT, a silencing mediator for retinoid and thyroid hormone
RT receptors-extended isoform that is more related to the nuclear
RT receptor corepressor.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:3519-3524(1999).
CC -!- FUNCTION: MEDIATES THE TRANSCRIPTIONAL REPRESSION ACTIVITY OF SOME
CC NUCLEAR RECEPTORS BY PROMOTING CHROMATIN CONDENSATION, THUS
CC PREVENTING ACCESS OF THE BASAL TRANSCRIPTION.
CC -!- SUBUNIT: FORMS A LARGE COREPRESSOR COMPLEX THAT CONTAINS SIN3A/B
CC AND HISTONE DEACETYLASES HDAC1 AND HDAC2. THIS COMPLEX ASSOCIATES
CC WITH THE THYROID (TR) AND THE RETINOIC ACID RECEPTORS (RAR) IN THE
CC ABSENCE OF LIGAND, AND MAY STABILIZE THEIR INTERACTION WITH TFIIIB.
CC -!- SUBCELLULAR LOCATION: NUCLEAR.
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; ALPHA (SHOWN HERE) AND BETA; ARE
CC PRODUCED BY ALTERNATIVE SPLICING.
CC -!- TISSUE SPECIFICITY: UBIQUITOUS. ALSO WIDELY EXPRESSED IN EARLY
CC EMBRYOS.
CC -!- DOMAIN: THE N-TERMINAL REGION CONTAINS REPRESSION FUNCTIONS THAT
CC ARE DIVIDED INTO THREE INDEPENDANT REPRESSION DOMAINS (RD1, RD2
CC AND RD3). THE C-TERMINAL REGION CONTAINS THE NUCLEAR RECEPTOR-
CC INTERACTING DOMAINS THAT ARE DIVIDED IN TWO SEPARATE INTERACTION
CC DOMAINS (ID1 AND ID2).
CC -!- DOMAIN: THE TWO INTERACTION DOMAINS (ID) CONTAIN A CONSERVED
CC SEQUENCE REFERRED TO AS THE CORNR BOX. THIS MOTIF IS REQUIRED AND
CC SUFFICIENT TO PERMIT BINDING TO UNLIGANDED TR AND RARS. SEQUENCES
CC FLANKING THE CORNR BOX DETERMINE NUCLEAR HORMONE RECEPTOR
CC SPECIFICITY.
CC -!- SIMILARITY: CONTAINS 1 SANT DOMAIN (COMPOSED OF SANT-A+SANT-B).
CC -!- SIMILARITY: CONTAINS 2 CORNR BOX.
CC -!- SIMILARITY: BELONGS TO THE N-COR NUCLEAR RECEPTOR COREPRESSORS
CC FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
```

```
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
CC EMBL; AF113001; RAD20944.1; -
CC EMBL; AF113002; AAD20945.1; -
CC EMBL; AF125671; AAD22972.1; -
CC MGD; MGI:1337080; Ncor2.
CC InterPro; IPR001005; Myb_DNA_bind.
CC Pfam; PF00249; myb_DNA-binding; 2.
CC SMART; SM00395; SANT; 2.
CC PROSITE; PS50090; MYB_3; 1.
KW Nuclear protein; Transcription regulation; DNA-binding; Repressor;
KW Coiled coil; Alternative splicing.
FT DOMAIN 165 207
FT DNA_BIND 254 312
FT SANT-A (POTENTIAL).
FT DNA_BIND 429 474
FT SANT-B (POTENTIAL).
FT DOMAIN 608 653
FT COILED COIL (POTENTIAL).
FT DOMAIN 652 682
FT COILED COIL (POTENTIAL).
FT DOMAIN 775 804
FT PRO-RICH.
FT DOMAIN 989 999
FT PRO-RICH.
FT DOMAIN 1351 1357
FT PRO-RICH.
FT DOMAIN 2094 2098
FT CORNR BOX OF ID1.
FT DOMAIN 2296 2300
FT CORNR BOX OF ID2.
FT DOMAIN 494 507
FT POLY-GLN.
FT DOMAIN 1615 1619
FT POLY-ALA.
FT DOMAIN 2434 2437
FT POLY-PRO.
FT VARSPLIC 36 254
FT MISSING (IN ISOFORM BETA).
FT CONFLICT 176 176
FT CONFLICT 396 402
FT CONFLICT 555 555
FT CONFLICT 756 756
FT CONFLICT 785 785
FT CONFLICT 806 846
FT CONFLICT 856 856
FT CONFLICT 859 859
FT CONFLICT 867 867
FT CONFLICT 895 895
FT CONFLICT 916 916
FT CONFLICT 975 975
FT CONFLICT 1046 1063
FT CONFLICT 1073 1080
FT CONFLICT 1133 1133
FT CONFLICT 1149 1149
FT CONFLICT 1157 1157
FT CONFLICT 1172 1201
FT CONFLICT 1696 1696
FT CONFLICT 1855 1857
FT CONFLICT 1909 1909
FT CONFLICT 1913 1913
FT CONFLICT 1923 1923
FT CONFLICT 1956 1956
FT CONFLICT 1968 1968
FT CONFLICT 2195 2196
FT CONFLICT 2213 2214
FT CONFLICT 2224 2224
FT CONFLICT 2472 2472
FT SEQUENCE 2472 AA; 270856 MW; 2A58F4DF7B79285B CRC64;
QY 85 PDHGT 91
DB 2143 PDHGT 2149
```

```
Query Match 5.7%; Score 7; DB 1; Length 2472;
Best Local Similarity 100.0%; Pred. No. 25;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 85 PDHGT 91
DB 2143 PDHGT 2149
```

```
RESULT 7
YVGB_VACCC STANDARD; PRT; 91 AA.
AC P20547;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL 11.2 KDA PROTEIN.
GN B ORF G AND C ORF F.
OS Vaccinia virus (strain Copenhagen).
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Orthopoxvirus.
OX NCBI_TaxID=10249;
RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE=91021027; PubMed=2219722;
RA Goebel S.J., Johnson G.P., Perkus M.E., Davis S.W., Winslow J.P.,
RA Paoletti E.;
RT "The complete DNA sequence of vaccinia virus.";
RL Virology 179:247-266(1990).
RN [2]
RN COMPLETE GENOME.
RP Goebel S.J., Johnson G.P., Perkus M.E., Davis S.W., Winslow J.P.,
RA Paoletti E.;
RT "Appendix to 'The complete DNA sequence of vaccinia virus'.";
RL Virology 179:517-563(1990).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M35027; AAA48224.1; -.
CC EMBL; M35027; AAA47977.1; -.
CC PIR; C33172; C33172.
CC PIR; H42529; H42529.
KW Hypothetical protein.
SQ SEQUENCE 91 AA; 11195 MW; 3621A57A685C9421 CRC64;
Query Match 4.9%; Score 6; DB 1; Length 91;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 98 KYRVYV 103
Db |||||
31 KYRVYV 36
RESULT 8
PT27_STYPL STANDARD; PRT; 114 AA.
AC P28219;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 01-DEC-1992 (Rel. 24, Last annotation update)
DE PROTEIN-TYROSINE PHOSPHATASE 27 (EC 3.1.3.48) (FRAGMENT).
GN STY 27.
OS Styela plicata (Sea squirt).
OC Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea;
OC Stolidobranchia; Styelidae; Styela.
OX NCBI_TaxID=7726;
RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE=91139172; PubMed=1704870;
RA Matthews R.J., Flores E., Thomas M.L.;
RT "Protein tyrosine phosphatase domains from the protochordate Styela
RT plicata.";
RL Immunogenetics 33:33-41(1991).
CC -!- CATALYTIC ACTIVITY: PROTEIN TYROSINE PHOSPHATE + H(2)O =
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M38010; AAA29843.1; -.
CC HSSP; P18052; 1YFO.
DR InferPro; IPR000340; DS_phosphatase.
DR InferPro; IPR000387; TYR_phosphatase.
DR InferPro; IPR000242; TYR_prot_phptase.
DR Pfam; PF00102; Y_phosphatase; 1.
DR SMART; SM00012; PTPC_DSPC; 1.
RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE=91139172; PubMed=1704870;
RA Matthews R.J., Flores E., Thomas M.L.;
RT "Protein tyrosine phosphatase domains from the protochordate Styela
RT plicata.";
RL Immunogenetics 33:33-41(1991).
CC -!- CATALYTIC ACTIVITY: PROTEIN TYROSINE PHOSPHATE + H(2)O =
```

```
CC -----
CC PROTEIN TYROSINE + ORTHOPHOSPHATE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M38012; AAA29819.1; -.
CC HSSP; P28827; 1REP.
DR InferPro; IPR000340; DS_phosphatase.
DR InferPro; IPR000387; TYR_phosphatase.
DR InferPro; IPR000242; TYR_prot_phptase.
DR Pfam; PF00102; Y_phosphatase; 1.
DR SMART; SM00012; PTPC_DSPC; 1.
DR PROSITE; PS00383; TYR_PHOSPHATASE_1; PARTIAL.
DR PROSITE; PS00056; TYR_PHOSPHATASE_2; PARTIAL.
DR PROSITE; PS00055; TYR_PHOSPHATASE_PTP; 1.
KW Hydrolase.
FT NON_TER 1
FT NON_TER 114
SQ SEQUENCE 114 AA; 13375 MW; 80D335FA4C72CFA1 CRC64;
Query Match 4.9%; Score 6; DB 1; Length 114;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 83 KWPDPHG 88
Db |||||
79 KWPDPHG 84
RESULT 9
PT25_STYPL STANDARD; PRT; 117 AA.
AC P28217;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 01-DEC-1992 (Rel. 24, Last annotation update)
DE PROTEIN-TYROSINE PHOSPHATASE 25 (EC 3.1.3.48) (FRAGMENT).
GN STY 25.
OS Styela plicata (Sea squirt).
OC Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea;
OC Stolidobranchia; Styelidae; Styela.
OX NCBI_TaxID=7726;
RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE=91139172; PubMed=1704870;
RA Matthews R.J., Flores E., Thomas M.L.;
RT "Protein tyrosine phosphatase domains from the protochordate Styela
RT plicata.";
RL Immunogenetics 33:33-41(1991).
CC -!- CATALYTIC ACTIVITY: PROTEIN TYROSINE PHOSPHATE + H(2)O =
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M38010; AAA29843.1; -.
CC HSSP; P18052; 1YFO.
DR InferPro; IPR000340; DS_phosphatase.
DR InferPro; IPR000387; TYR_phosphatase.
DR InferPro; IPR000242; TYR_prot_phptase.
DR Pfam; PF00102; Y_phosphatase; 1.
DR SMART; SM00012; PTPC_DSPC; 1.
```

DR PROSITE; PS00383; TYR\_PHOSPHATASE\_1; PARTIAL.  
 DR PROSITE; PS00056; TYR\_PHOSPHATASE\_2; PARTIAL.  
 DR PROSITE; PS00055; TYR\_PHOSPHATASE\_PTP; 1.  
 KW Hydrolase.  
 FT NON\_TER 1  
 FT NON\_TER 117  
 FT NON\_TER 117  
 SQ SEQUENCE 117 AA; 13422 MW; 9CAE0F5D70B7EF28 CRC64;

Query Match 4.9%; Score 6; DB 1; Length 117;  
 Best Local Similarity 100.0%; Pred. No. 21;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 83 KWPDHG 88  
 | | | | |  
 DB 82 KWPDHG 87

RESULT 10  
 YAE6\_YEAST  
 ID YAE6\_YEAST STANDARD; PRT; 118 AA.  
 AC P39724;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE HYPOTHETICAL 13.4 KDA PROTEIN IN ACS1-GCV3 INTERGENIC REGION.  
 GN YAL046C.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288C / AB972;  
 RX MEDLINE=95249563; PubMed=7731988;  
 RA Bussey H., Kaback D.B., Zhong W., Vo D.T., Clark M.W., Fortin N.,  
 RA Hall J., Ouellette B.F.F., Keng T., Barton A.B., Su Y., Davies C.K.,  
 RA Storms R.K.;  
 RT "The nucleotide sequence of chromosome I from Saccharomyces  
 RT cerevisiae.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 92:3809-3813(1995).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; U12980; AAC04985.1; -  
 DR SGD; S0000044; YAL046C.  
 DR InterPro; IPR002634; BOLA.  
 DR Pfam; PF01722; BOLA; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 118 AA; 13356 MW; 244F5FF2052FF410 CRC64;

Query Match 4.9%; Score 6; DB 1; Length 118;  
 Best Local Similarity 100.0%; Pred. No. 21;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 67 KKSTGK 72  
 | | | | |  
 DB 108 KKSTGK 113

RESULT 11  
 RL31\_SULAC  
 ID RL31\_SULAC STANDARD; PRT; 129 AA.  
 AC P38618;  
 DT 01-OCT-1994 (Rel. 30, Created)  
 DT 01-OCT-1994 (Rel. 30, Last sequence update)

DT 20-AUG-2001 (Rel. 40, Last annotation update)  
 DE 50S RIBOSOMAL PROTEIN L31E.  
 GN RPL31E.  
 OS Sulfolobus acidocaldarius.  
 OC Archaea; Crenarchaeota; Sulfolobales; Sulfolobaceae; Sulfolobus.  
 OX NCBI\_TaxID=2285;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 33909 / NCIB 11770 / DSM 639;  
 RX MEDLINE=95226466; PubMed=7711082;  
 RA Moll R., Schaefer G., Schmidtke S.;  
 RT "Nucleotide sequence of a gene cluster encoding ribosomal proteins in  
 RT the thermoacidophilic crenarchaeon Sulfolobus acidocaldarius.";  
 RL Bicchim. Biophys. Acta 1261:315-318(1995).  
 CC -!- SIMILARITY: BELONGS TO THE L31E FAMILY OF RIBOSOMAL PROTEINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; X77509; CAA54639.1; -  
 DR F1R; S41964; S41964.  
 DR InterPro; IPR000054; Ribosomal\_L31e.  
 DR ProbDom; PD006030; Ribosomal\_L31e; 1.  
 DR PROSITE; PS01144; RIBOSOMAL\_L31E; FALSE\_NEG.  
 KW Ribosomal protein.  
 SQ SEQUENCE 129 AA; 15196 MW; 56ACF37C68E076A0 CRC64;

Query Match 4.9%; Score 6; DB 1; Length 129;  
 Best Local Similarity 100.0%; Pred. No. 23;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 97 IKYVRY 102  
 | | | | |  
 DB 70 IKYVRY 75

RESULT 12  
 CSF2\_FELCA  
 ID CSF2\_FELCA STANDARD; PRT; 144 AA.  
 AC O62757;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE GRANULOCYTE-MACROPHAGE COLONY-STIMULATING FACTOR PRECURSOR (GM-CSF)  
 DE (COLONY-STIMULATING FACTOR) (CSF).  
 GN CSF2.  
 OS Felis silvestris catus (Cat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.  
 OX NCBI\_TaxID=9685;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Hughes K.J., O'Reilly K.L.;  
 RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.  
 CC -!- FUNCTION: CYTOKINE THAT STIMULATES THE GROWTH AND DIFFERENTIATION  
 CC OF HEMATOPOIETIC PRECURSOR CELLS FROM VARIOUS LINEAGES, INCLUDING  
 CC GRANULOCYTES, MACROPHAGES, EOSINOPHILS AND ERYTHROCYTES (BY  
 CC SIMILARITY).  
 CC -!- SUBUNIT: MONOMER (BY SIMILARITY).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----



```
CC EMBL; AF053007; AAC06041.1; -.
DR InterPro: IPR000773; GM_CSF.
DR Pfam: PF01109; GM_CSF; 1.
DR PRINTS; PR00693; GMCSEFACTOR.
DR PRODom; PD007349; GM_CSF; 1.
DR SMART; SM00040; CSF2; 1.
DR PROSITE; PS00702; GM_CSF; 1.
KW Cytokine; Growth factor; Glycoprotein; Signal.
FT SIGNAL 1 17
FT CHAIN 18 144
FT
FT DISULFID 71 113
FT BY SIMILARITY.
FT DISULFID 105 138
FT BY SIMILARITY.
FT CARBOHYD 44 44
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 45 45
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 54 54
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 144 AA; 16505 MW; 42C5BF5F8235DA55 CRC64;

Query Match 4.9%; Score 6; DB 1; Length 144;
Best Local Similarity 100.0%; Pred. No. 26;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 35 SLKEPL 40
DB 89 SLKEPL 94

RESULT 13
SP1A_STRPU STANDARD; PRT; 152 AA.
ID SP1A_STRPU STANDARD; PRT; 152 AA.
AC P04109; 1986 (Rel. 03, Created)
DT 01-NOV-1986 (Rel. 03, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE SPEC 1A PROTEIN.
GN SPC1.
OS Strongylocentrotus purpuratus (Purple sea urchin).
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC Echinoidea; Euechinozoa; Echinacea; Echinozoa; Strongylocentrotidae;
OC Strongylocentrotus.
OX NCBI_TaxID=7668;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Gastrula;
RX MEDLINE=88011349; PubMed=3116272;
RA Hardin P.E., Klein W.H.;
RT "Unusual sequence conservation in the 5' and 3' untranslated regions
of the sea urchin spec mRNAs.";
RL J. Mol. Evol. 25:126-133(1987).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=86115267; PubMed=2935638;
RA Hardin S.H., Carpenter C.D., Hardin P.E., Bruskin A.M., Klein W.H.;
RT "Structure of the Sperl gene encoding a major calcium-binding protein
in the embryonic ectoderm of the sea urchin, Strongylocentrotus
purpuratus.";
RL J. Mol. Biol. 186:243-255(1985).
RN [3]
RP SEQUENCE OF 67-152 FROM N.A.
RX MEDLINE=84130177; PubMed=6697391;
RA Carpenter C.D., Bruskin A.M., Hardin P.E., Keast M.J., Anstrom J.A.,
RA Tyner A.L., Brandhorst B.P., Klein W.H.;
RT "Novel proteins belonging to the troponin C superfamily are encoded
by a set of mRNAs in sea urchin embryos.";
RL Cell 36:663-671(1984).
CC -1- FUNCTION: CALCIUM-BINDING PROTEIN INVOLVED IN LARVAL DEVELOPMENT
AND METAMORPHOSIS. LIKELY TO FUNCTION AS CALCIUM BUFFERS
CC MEDIATING THE TRANSPORT OF CALCIUM FROM THE SEA WATER TO THE
CC BLASTOCOELE WHERE CALCIUM IS REQUIRED FOR SKELETON FORMATION.
CC -1- TISSUE SPECIFICITY: FOUND IN CELL LINEAGES GIVING RISE TO THE
ABORAL ECTODERM, A SQUAMOUS EPITHELIUM COVERING THE SURFACE OF THE
```

```
CC LATE STAGE EMBRYO AND LARVA.
CC -1- DEVELOPMENTAL STAGE: ACCUMULATE IN EMBRYOS AND LARVAE, BUT NOT IN
CC ADULTS.
CC -1- SIMILARITY: TO OTHER EF-HAND CALCIUM BINDING PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; U38804; AAC08266.1; -  
DR Mendel; 10371; PORpu; ycf21.1.  
KW Chloroplast; Hypothetical protein.

SQ SEQUENCE 174 AA; 20632 MW; 7FF9D65664994E0E CRC64;

#### Query Match

Best Local Similarity 4.9%; Score 6; DB 1; Length 174;  
Matches 6; Conservative 100.0%; Pred. No. 30;

Mismatches 0; Indels 0; Gaps 0;

QY 48 LLENFQ 53

|||||

DB 169 LLENFQ 174

#### RESULT 15

SLYD\_HELPJ STANDARD; PRT; 185 AA.

AC Q9ZR89;

DT 20-AUG-2001 (Rel. 40, Created)

DT 20-AUG-2001 (Rel. 40, Last sequence update)

DT 20-AUG-2001 (Rel. 40, Last annotation update)

DE FKBP-TYPE PEPTIDYL-PROLYL CIS-TRANS ISOMERASE SLYD (EC 5.2.1.8)

DE (PPIASE) (ROTAMASE).

GN SLYD OR JHPI052.

OS Helicobacter pylori J99 (Campylobacter pylori J99).

OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;

OC Helicobacter.

OX NCBI\_TaxID=85963;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=99120557; PubMed=9923682;

RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,

RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,

RA Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,

RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,

RA Trust T.J.;

RT "Genomic sequence comparison of two unrelated isolates of the human

RT gastric pathogen Helicobacter pylori.";

RL Nature 397:176-180(1999).

CC -!- FUNCTION: PPIASES ACCELERATE THE FOLDING OF PROTEINS (BY

CC SIMILARITY).

CC -!- CATALYTIC ACTIVITY: CIS-TRANS ISOMERIZATION OF PROLINE IMIDIC

CC PEPTIDE BONDS IN OLIGOPEPTIDES.

CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (POTENTIAL).

CC -!- SIMILARITY: BELONGS TO THE FKBP-TYPE PPIASE FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

CC EMBL; AE001533; RAD06631.1; -

DR InterPro; IPR001179; FKBP\_PPIase.

DR Pfam; PF00254; FKBP; 1.

DR PROSITE; PS00453; FKBP\_PPIASE\_1; FALSE\_NEG.

DR PROSITE; PS00454; FKBP\_PPIASE\_2; FALSE\_NEG.

DR PROSITE; PS00559; FKBP\_PPIASE\_3; 1.

KW Isomerase; Rotamase; Complete proteome.

FT DOMAIN 1 99 PPIASE, FKBP-TYPE.

FT DOMAIN 158 185 GLY-RICH.

FT SEQUENCE 185 AA; 20100 MW; 2D48E82AC450B0CB CRC64;

#### Query Match

Best Local Similarity 4.9%; Score 6; DB 1; Length 185;  
Matches 6; Conservative 100.0%; Pred. No. 32;

Mismatches 0; Indels 0; Gaps 0;

QY 37 KEPLF 42

DB 34 KEPLF 39

Search completed: March 30, 2002, 08:31:33

Job time: 254 sec.

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 30, 2002, 08:26:48 ; Search time 27.01 Seconds  
(without alignments)  
660.690 Million cell updates/sec

Title: US-09-095-478A-8  
Perfect score: 122  
Sequence: 1 DFQGMWENNCNVIAMITRE.....VRKSHITGLLVHCTAGVGR 122

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 473505 seqs, 146272329 residues

Word size : 0

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

- Database : SPTREMBL17.\*
- 1: sp\_archaea.\*
  - 2: sp\_bacteria.\*
  - 3: sp\_fungi.\*
  - 4: sp\_human.\*
  - 5: sp\_invertebrate.\*
  - 6: sp\_mammal.\*
  - 7: sp\_mhc.\*
  - 8: sp\_organelle.\*
  - 9: sp\_phase.\*
  - 10: sp\_plant.\*
  - 11: sp\_rodent.\*
  - 12: sp\_virus.\*
  - 13: sp\_vertebrate.\*
  - 14: sp\_unclassified.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	23	18.9	426	11 O55082	O55082 mus musculu
2	10	8.2	398	4 Q9Y406	Q9Y406 homo sapien
3	7	5.7	52	7 Q9TPH8	Q9TPH8 xiphophorus
4	7	5.7	116	11 Q9D792	Q9D792 mus musculu
5	5	5.7	158	12 Q73103	Q73103 human immun
6	7	5.7	244	7 Q62884	Q62884 xiphophorus
7	7	5.7	310	5 Q9XV30	Q9XV30 caenorhabdi
8	7	5.7	351	10 Q9SST9	Q9SST9 scopharia du
9	7	5.7	378	10 Q9M7Q0	Q9M7Q0 porphyra pu
10	7	5.7	378	10 Q9M7P9	Q9M7P9 porphyra pu
11	7	5.7	378	10 Q9M7P8	Q9M7P8 porphyra pu
12	7	5.7	397	5 Q9XV85	Q9XV85 caenorhabdi
13	7	5.7	407	2 Q9AA59	Q9AA59 caulobacter
14	7	5.7	446	5 Q9Y1X4	Q9Y1X4 ephydatia f
15	7	5.7	488	5 Q9NL12	Q9NL12 branchiosto
16	7	5.7	542	5 Q24198	Q24198 drosophila
17	7	5.7	542	5 Q9VIB9	Q9VIB9 drosophila
18	7	5.7	991	2 Q9L773	Q9L773 xanthomonas
19	7	5.7	1214	5 Q9VAL3	Q9VAL3 drosophila

20	7	5.7	1286	5 Q9TZM3	Q9TZM3 caenorhabdi
21	6	4.9	41	2 Q31883	Q31883 bacillus su
22	6	4.9	41	9 Q64165	Q64165 bacterioph
23	6	4.9	48	2 Q48270	Q48270 helicobacte
24	6	4.9	49	2 Q9JX85	Q9JX85 neisseria m
25	6	4.9	91	12 Q78450	Q78450 human immun
26	6	4.9	92	12 Q80035	Q80035 human immun
27	6	4.9	94	2 Q50759	Q50759 borrelia bu
28	6	4.9	98	1 Q9HPX8	Q9HPX8 halobacteri
29	6	4.9	115	2 Q9RD28	Q9RD28 streptomyce
30	6	4.9	116	1 Q58282	Q58282 pyrococcus
31	6	4.9	116	2 Q32470	Q32470 aeromonas p
32	6	4.9	120	5 Q9U9U6	Q9U9U6 cryptospori
33	6	4.9	120	5 Q9TW73	Q9TW73 cryptospori
34	6	4.9	120	5 Q9TW72	Q9TW72 cryptospori
35	6	4.9	120	5 Q9TW71	Q9TW71 cryptospori
36	6	4.9	120	5 Q9TVU9	Q9TVU9 cryptospori
37	6	4.9	120	5 Q9NA06	Q9NA06 drosophila
38	6	4.9	120	5 Q9NA05	Q9NA05 drosophila
39	6	4.9	120	5 Q9NA04	Q9NA04 drosophila
40	6	4.9	120	5 Q9N9Y2	Q9N9Y2 drosophila
41	6	4.9	120	5 Q9N9Y1	Q9N9Y1 drosophila
42	6	4.9	120	5 Q9N9Y0	Q9N9Y0 drosophila
43	6	4.9	120	5 Q9N6H0	Q9N6H0 drosophila
44	6	4.9	124	7 Q08610	Q08610 aulonocara
45	6	4.9	142	11 Q9D7H4	Q9D7H4 mus musculu

ALIGNMENTS

RESULT 1

O55082 ID O55082 PRELIMINARY; PRT; 426 AA.

AC O55082;

DT 01-JUN-1998 (Tremblrel. 06, Created)

DT 01-JUN-1998 (Tremblrel. 06, Last sequence update)

DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)

DE PROTEIN TYROSINE PHOSPHATASE, NON-RECEPTOR TYPE 20 (EC 3.1.3.48)

DE (PHOSPHOTYROSINE PHOSPHATASE) (PTPASE).

GN PTPN20.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.

OX NCPI\_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=TESTIS;

RA MEDLINE=98070510; Pubmed=9407093;

RA Ohsugi M., Kuramochi S., Matsuda S., Yamamoto T.;

RT Molecular cloning and characterization of a novel cytoplasmic

RT protein-tyrosine phosphatase that is specifically expressed in

RT spermatocytes.";

RL J. Biol. Chem. 272:33092-33099(1997).

CC -; CATALYTIC ACTIVITY: PROTEIN TYROSINE PHOSPHATE + H(2)O = PROTEIN

CC -; TYROSINE + PHOSPHATE.

DR EMBL; D64141; BAA23761.1; -.

DR HSSP; P28827; IRPM.

DR MGD; MGI:1196295; Ptpn20.

DR InterPro; IPR000387; Tyr\_phosphatase.

DR InterPro; IPR000242; Tyr\_prot\_phptase.

DR Pfam; PF00102; Y\_phosphatase; 1.

DR PRINTS; PR00700; PRTYPHPTASE.

 DR SMART; SM00194; PTPC; 1. | DR PROSITE; PS00383; TYR\_PHOSPHATASE\_1; 1. | DR PROSITE; PS00056; TYR\_PHOSPHATASE\_2; 1. | DR PROSITE; PS00055; TYR\_PHOSPHATASE\_PTP; 1. | KW Hydrolase. | SQ SEQUENCE 426 AA; 49118 MW; 2B35FB13379502F4 CRC64; |

Query Match 18.9%; Score 23; DB 11; Length 426;  
Best Local Similarity 100.0%; Pred. No. 5.7e-17;

```

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 94 DFFIKYRVYRKSHITGPLLVC 116
    |||||
Db 337 DFFIKYRVYRKSHITGPLLVC 359

RESULT 2
QY406
ID QY406 PRELIMINARY; PRT; 398 AA.
AC QY406;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE HYPOTHETICAL 45.7 KDA PROTEIN (EC 3.1.3.48) (FRAGMENT).
GS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=KIDNEY;
RA Anorge W., Winkler U., Mewes H.W., Gassenhuber J., Wiemann S.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL050040; CAB43248.1; -.
DR HSSP; Q06124; 2SHP.
DR InterPro; IPR000387; TYR_phosphatase.
DR Pfam; PF00102; Y_phosphatase; 1.
DR PRINTS; PR00700; PRTYPHPTASE.
DR SMART; SM00194; PTPC; 1.
DR PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
DR PROSITE; PS00556; TYR_PHOSPHATASE_2; 1.
DR PROSITE; PS00555; TYR_PHOSPHATASE_PTP; 1.
KW Hydrolase; Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 398 AA; 45690 MW; 857AAD03747870A2 CRC64;

Query Match
Best Local Similarity 8.2%; Score 10; DB 4; Length 398;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 83 KWPDHGTPAS 92
    |||||
Db 298 KWPDHGTPAS 307

RESULT 3
QYTPH8
ID QYTPH8 PRELIMINARY; PRT; 52 AA.
AC QYTPH8;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)
DE MHC CLASS II ANTIGEN (FRAGMENT).
OS Xiphophorus helleri.
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percormorpha; Atherinomorpha;
OC Cyprinodontiformes; Poeciliidae; Xiphophorus.
OX NCBI_TaxID=8084;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MHCIIASWORDTAIL;
RX MEDLINE=99398697; PubMed=10468597;
RA Venkatesh B., Ning Y., Brenner S.;
RT "Late changes in spliceosomal introns define clades in vertebrate
evolution.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:10267-10271(1999).
DR EMBL; AF134946; AAD54146.1; -.
KW MHC.

```

```

FT NON_TER 1
SQ SEQUENCE 52 AA; 5839 MW; 07CFF4E11374DA1B CRC64;

Query Match
Best Local Similarity 5.7%; Score 7; DB 7; Length 52;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 VSLKEPL 40
    |||||
Db 23 VSLKEPL 29

RESULT 4
QYD792
ID QYD792 PRELIMINARY; PRT; 116 AA.
AC QYD792;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE 913001LK15RIK PROTEIN.
GN 913001LK15RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=TONGUE;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Sakimi L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK009447; BAB26293.1; -.
DR MGI; MGI:1914070; 913001LK15RIK.
SQ SEQUENCE 116 AA; 13274 MW; FE28999B4A2B25EA8 CRC64;

Query Match
Best Local Similarity 5.7%; Score 7; DB 11; Length 116;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 22 EGGVIKC 28
    |||||
Db 79 EGGVIKC 85

RESULT 5
QY3103
ID QY3103 PRELIMINARY; PRT; 158 AA.
AC QY3103;
DT 01-AUG-1998 (TREMBlrel. 07, Created)
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE ENVELOPE GLYCOPROTEIN V3-V5 LOOP REGION (FRAGMENT).

```

OS Human immunodeficiency virus type 1.  
 OC Viruses; Retroviral viruses; Retroviridae; Lentivirus.  
 OX NCBI\_TaxID=11676;

RN [1]  
 RP SEQUENCE FROM N.A.

RX MEDLINE=98160778; PubMed=9499799;  
 RA Shankarappa R., Gupta P., Learn G.H. Jr., Rodrigo A.G., Ehrlich G.D.;  
 RA Rinaldo C.R. Jr., Gorro M.C., Millins J.I., Nara P.L., Ehrlich G.D.;  
 RT "Evolution of human immunodeficiency virus type 1 envelope sequences  
 in infected individuals with differing disease progression profiles."  
 RL Virology 241:251-259(1998).

DR EMBL: U15836; AAC59204.1; -.

DR InterPro: IPR000777; GP120.

DR Pfam: PF00516; GP120; 1.

KW Envelope protein.

FT NON\_TER 1

FT NON\_TER 158

SQ SEQUENCE 158 AA; 17431 MW; 7039619AB3EB8595 CRC64;

Query Match 5.7%; Score 7; DB 12; Length 158;

Best Local Similarity 100.0%; Pred. No. 12;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 108 ITGPLLIV 114

DB 142 ITGPLLIV 148

RESULT 6

ID O62884 PRELIMINARY; PRT; 244 AA.

AC O62884;

DT 01-AUG-1998 (TREMBlrel. 07, Created)

DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)

DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)

DE MHC CLASS II BETA CHAIN (FRAGMENT).

GN XIMA-DAB.

OS Xiphophorus maculatus (Southern platyfish).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;

OC Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;

OC Cyprinodontiformes; Poeciliidae; Xiphophorus.

OX NCBI\_TaxID=8083;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=JP 163 A; TISSUE=GUT;

RA McConnell T.J., Godwin U.B., Norton S.F., Nairn R.S., Kazianis S.,

RA Morizot D.C.;

RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.

CC -!- MISCELLANEOUS: FOUND N-TERMINAL TO IMMUNOGLOBULIN AND MAJOR

CC HISTOCOMPATIBILITY COMPLEX DOMAIN (BY SIMILARITY).

CC -!- SIMILARITY: TO IMMUNOGLOBULIN AND MAJOR HISTOCOMPATIBILITY COMPLEX

CC DOMAIN.

CC -!- SIMILARITY: TO CLASS II HISTOCOMPATIBILITY ANTIGEN, BETA CHAIN,

CC BETA-1 DOMAIN INTERPRO FAMILY.

DR EMBL: AF040760; AAC05652.1; -.

DR HSSP: P01888; IBMG.

DR InterPro: IPR003597; Ig\_c1.

DR InterPro: IPR003006; Ig\_MHC.

DR InterPro: IPR000353; MHC\_II\_beta.

DR Pfam: PF00047; Ig; 1.

DR Pfam: PF00969; MHC\_II\_beta; 1.

DR ProDom: PD000328; MHC\_II\_beta; 1.

DR SMART: SM00407; IGcl; 1.

KW Glycoprotein; MHC; MHC II; Transmembrane.

FT NON\_TER 1

FT NON\_TER 244

SQ SEQUENCE 244 AA; 27861 MW; 8B0BDC0224039580 CRC64;

Query Match

Best Local Similarity 100.0%; Score 7; DB 7; Length 244;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 34 VSLKEPL 40  
 DB 188 VSLKEPL 194

RESULT 7

ID Q9XV30 PRELIMINARY; PRT; 310 AA.

AC Q9XV30;

DT 01-NOV-1999 (TREMBlrel. 12, Created)

DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)

DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)

DE F40D4.5 PROTEIN.

GN F40D4.5.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;

OC Rhabditidae; Peloderinae; Caenorhabditis.

OX NCBI\_TaxID=6239;

RN [1]

RP SEQUENCE FROM N.A.

RA Matthews L.;

RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=94150718; PubMed=7906398;

RA Wilson R., Almscough R., Anderson K., Baynes C., Berks M.,

RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,

RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,

RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,

RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,

RA Lightning J., Lloyd C., Murray A., Mortimore B., O'Callaghan M.,

RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,

RA Smaildon N., Smith A., Sonhammer E., Staden R., Sulston J.,

RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,

RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;

RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.

RT elegans."

RL Nature 368:32-38(1994).

DR EMBL: Z81536; CAB04364.1; -.

DR InterPro: IPR000276; GPCR\_Rhodpsn.

DR PROSITE: PS50262; G\_PROTEIN\_RECEP\_F1\_2; 1.

SQ SEQUENCE 310 AA; 35478 MW; FA6A377FD2A76C0C CRC64;

Query Match

Best Local Similarity 100.0%; Score 7; DB 5; Length 310;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 108 ITGPLLIV 114

DB 234 ITGPLLIV 240

RESULT 8

ID Q9SST9 PRELIMINARY; PRT; 351 AA.

AC Q9SST9;

DT 01-MAY-2000 (TREMBlrel. 13, Created)

DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)

DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)

DE GERANYLGERANYL PYROPHOSPHATE SYNTHASE.

OS Scoparia dulcis.

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;

OC Asteridae; euasterids I; Lamiales; Veronicaceae; Scoparia.

OX NCBI\_TaxID=107240;

RN [1]

RP SEQUENCE FROM N.A.

RA Kojima N., Suh D., Sankawa U.;

RT "Scoparia dulcis GGPP synthase."

RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL: AB034250; BAA86285.1; -.

DR InterPro: IPR000092; Polyprenyl\_synt.  
 DR Pfam: PF00348; polyprenyl\_synt; 1.  
 DR PROSITE; PS00723; POLYPRENYL\_SYNTHET\_1; 1.  
 DR PROSITE; PS00444; POLYPRENYL\_SYNTHET\_2; 1.  
 SQ SEQUENCE 351 AA; 38342 MW; 247570E3CD863D5C CRC64;

Query Match 5.7%; Score 7; DB 10; Length 351;  
 Best Local Similarity 100.0%; Pred. No. 23;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 34 VSLKEPL 40  
 |||||  
 Db 79 VSLKEPL 85

RESULT 9

Q9M7Q0 Q9M7Q0 PRELIMINARY; PRT; 378 AA.  
 AC Q9M7Q0;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
 DE REPLICASE.  
 OS Porphyra pulchra.  
 OC Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Porphyra.  
 OX NCBI\_TaxID=60925;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Moon D.A., Goff L.G.;  
 RT "Sequence signatures of flowering plant geminiviruses are found in  
 dsDNA plasmids of the evolutionarily ancient red alga Porphyra.";  
 RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF106326; AAF36422.1; -;  
 DR InterPro: IPR001191; Gemini\_AL1.  
 DR ProDom: PD000736; Gemini\_AL1; 1.  
 SQ SEQUENCE 378 AA; 43495 MW; 056A183EA635FDC1 CRC64;

Query Match 5.7%; Score 7; DB 10; Length 378;  
 Best Local Similarity 100.0%; Pred. No. 25;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 44 HFHVLE 50  
 |||||  
 Db 62 HFHVLE 68

RESULT 10

Q9M7P9 Q9M7P9 PRELIMINARY; PRT; 378 AA.  
 AC Q9M7P9;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
 DE REPLICASE.  
 OS Porphyra pulchra.  
 OC Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Porphyra.  
 OX NCBI\_TaxID=60925;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Moon D.A., Goff L.G.;  
 RT "Sequence signatures of flowering plant geminiviruses are found in  
 dsDNA plasmids of the evolutionarily ancient red alga Porphyra.";  
 RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF106327; AAF36423.1; -;  
 DR InterPro: IPR001191; Gemini\_AL1.  
 DR Pfam: PF00799; Gemini\_AL1; 1.  
 DR ProDom: PD000736; Gemini\_AL1; 1.  
 SQ SEQUENCE 378 AA; 43680 MW; 1C8E9E4A9E68EE14 CRC64;

Query Match 5.7%; Score 7; DB 10; Length 378;

Best Local Similarity 100.0%; Pred. No. 25;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 44 HFHVLE 50  
 |||||  
 Db 62 HFHVLE 68

RESULT 11

Q9M7P8 Q9M7P8 PRELIMINARY; PRT; 378 AA.  
 AC Q9M7P8;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
 DE REPLICASE.  
 OS Porphyra pulchra.  
 OC Eukaryota; Rhodophyta; Bangiophyceae; Bangiales; Porphyra.  
 OX NCBI\_TaxID=60925;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Moon D.A., Goff L.G.;  
 RT "Sequence signatures of flowering plant geminiviruses are found in  
 dsDNA plasmids of the evolutionarily ancient red alga Porphyra.";  
 RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF106328; AAF36424.1; -;  
 DR InterPro: IPR001191; Gemini\_AL1.  
 DR Pfam: PF00799; Gemini\_AL1; 1.  
 DR ProDom: PD000736; Gemini\_AL1; 1.  
 SQ SEQUENCE 378 AA; 44047 MW; FFE240CDBF4AEFE0 CRC64;

Query Match 5.7%; Score 7; DB 10; Length 378;  
 Best Local Similarity 100.0%; Pred. No. 25;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 44 HFHVLE 50  
 |||||  
 Db 62 HFHVLE 68

RESULT 12

Q9XV85 Q9XV85 PRELIMINARY; PRT; 397 AA.  
 AC Q9XV85;  
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)  
 DE F20E11.2 PROTEIN.  
 GN F20E11.2.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 CX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Ainscough R.;  
 RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94150718; PubMed=7906398;  
 RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,  
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,  
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,  
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,  
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,  
 RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,  
 RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,  
 RA Smalton N., Smith A., Sonnhammer E., Staden R., Sulston J.,  
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,  
 RA Watson A., Weinstock L., Wilkinon-Sproat J., Wohlman P.;  
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
 elegans.";

```
RL Nature 368:32-38(1994).
DR EMBL: Z81508; CAB04140.1; -.
DR InterPro: IPR000276; GPCR_Rhodpsn.
DR PROSITE: PS0262; G_PROTEIN_RECEP_F1_2; 1.
SQ SEQUENCE 397 AA; 44724 MW; 7BE8E2FFEE26EC4 CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 26; Length 397;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 108 ITGPLL 114
Db 234 ITGPLL 240
|||||

RESULT 13
Q9AA59 PRELIMINARY; PRT; 407 AA.
AC Q9AA59;
DT 01-JUN-2001 (TRENBLrel. 17, Created)
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
DE OMPA FAMILY PROTEIN.
GN CC0747.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC Caulobacter.
OX NCBI_TaxID=69394;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21173698; PubMed=11259647;
RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT "Complete genome sequence of Caulobacter crescentus."
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR EMBL: AE005750; AAK22732.1; -.
DR TIGR: CC0747; -.
KW Complete proteome.
SQ SEQUENCE 407 AA; 42781 MW; 54610D5FA64CC3BF CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 27; Length 407;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 116 CTAGVGR 122
Db 105 CTAGVGR 111
|||||

RESULT 14
Q9Y1X4 PRELIMINARY; PRT; 446 AA.
AC Q9Y1X4;
DT 01-NOV-1999 (TRENBLrel. 12, Created)
DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)
DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
DE SPTR5 (EC 3.1.3.48) (FRAGMENT).
OS Ephydatia fluviatilis.
OC Eukaryota; Metazoa; Porifera; Demospongiae; Ceractinomorpha;
OC Aplousclerida; Spongillidae; Ephydatia.
OX NCBI_TaxID=31330;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99246376; PubMed=10229569;
RA Ono K., Suga H., Iwabe N., Kuma K., Miyata T.;
RT "Multiple protein tyrosine phosphatases in sponges and explosive gene

RT duplication in the early evolution of animals before the parazoan-
emmetazoan split.";
J. Mol. Evol. 48:654-662(1999).
DR EMBL: AB019127; BAA82560.1; -.
DR HSSP: P18052; 1VFO.
DR InterPro: IPR000387; TYR_phosphatase.
DR InterPro: IPR000242; TYR_prot_phptase.
DR Pfam: PF00102; Y_phosphatase; 2.
DR PRINTS: PR00700; PRTYPHPTASE.
DR SMART: SM00194; PTPC; 2.
DR PROSITE: PS00383; TYR_PHOSPHATASE_1; 1.
DR PROSITE: PS00383; TYR_PHOSPHATASE_2; 2.
DR PROSITE: PS50056; TYR_PHOSPHATASE_2; 2.
DR PROSITE: PS50055; TYR_PHOSPHATASE_PTP; 2.
KW Hydrolase.
FT NON_TER 1
SQ SEQUENCE 446 AA; 50365 MW; 9E5B8AF6168FDF7C CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 29; Length 446;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 110 GPLLVHC 116
Db 152 GPLLVHC 108
|||||

RESULT 15
Q9NL12 PRELIMINARY; PRT; 488 AA.
ID Q9NL12;
AC Q9NL12;
DT 01-OCT-2000 (TRENBLrel. 15, Created)
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
DE AMPTPR5 PROTEIN (FRAGMENT).
GN AMPTPR5.
OS Branchiostoma belcheri.
OC Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;
OC Branchiostoma.
OX NCBI_TaxID=7741;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20219325; PubMed=10754074;
RA Ono-Koyanagi K., Suga H., Kato K., Miyata T.;
RT "Protein tyrosine phosphatases from amphioxus, hagfish, and ray:
RT divergence of tissue-specific isoform genes in the early evolution of
RT vertebrates.";
J. Mol. Evol. 50:302-311(2000).
DR EMBL: AB033565; BAA95172.1; -.
DR InterPro: IPR000387; TYR_phosphatase.
DR InterPro: IPR000242; TYR_prot_phptase.
DR Pfam: PF00102; Y_phosphatase; 2.
DR PRINTS: PR00700; PRTYPHPTASE.
DR SMART: SM00194; PTPC; 2.
DR PROSITE: PS00383; TYR_PHOSPHATASE_1; 1.
DR PROSITE: PS50056; TYR_PHOSPHATASE_2; 2.
DR PROSITE: PS50055; TYR_PHOSPHATASE_PTP; 2.
KW Hydrolase.
FT NON_TER 1
SQ SEQUENCE 488 AA; 55795 MW; A3DB8EB912DD7226 CRC64;

Query Match
Best Local Similarity 100.0%; Pred. No. 31; Length 488;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 84 WPDHGT 90
Db 78 WPDHGT 84
|||||

Search completed: March 30, 2002, 08:31:09
```

Job time: 261 sec

2 1 1 1



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2002, 04:20:42 ; Search time 1468.71 Seconds  
(without alignments)  
3594.377 Million cell updates/sec

Title: US-09-095-478A-4  
Perfect score: 320  
Sequence: 1 GAAATAATTGTAATGTTAT.....ATTACAGACCCCTCTGTGT 320

Scoring table:  
Gapop 10.0, Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

#### Database :

GenEmbl:\*

1: gb.ba.\*  
2: gb.htg.\*  
3: gb.in.\*  
4: gb.om.\*  
5: gb.ov.\*  
6: gb.pat.\*  
7: gb.ph.\*  
8: gb.pl.\*  
9: gb.pr.\*  
10: gb.ro.\*  
11: gb.sts.\*  
12: gb.sy.\*  
13: gb.un.\*  
14: gb.vi.\*  
15: em.ba.\*  
16: em.fun.\*  
17: em.hum.\*  
18: em.in.\*  
19: em.om.\*  
20: em.or.\*  
21: em.ov.\*  
22: em.pat.\*  
23: em.ph.\*  
24: em.pl.\*  
25: em.ro.\*  
26: em.sts.\*  
27: em.sy.\*  
28: em.un.\*  
29: em.vi.\*  
30: em.htgo.hum.\*  
31: em.htgo.inv.\*  
32: em.htgo.rod.\*  
33: em.htg.hum.\*  
34: em.htg.inv.\*  
35: em.htg.rod.\*  
36: em.htg.other.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

8

Result No.	Score	Match	Length	DB	ID	Description
1	280	87.5	3090	10	D64141	D64141 Mus musculus
2	241.6	75.5	2786	9	HS800374	AL050040 Homo sapi
3	144.8	45.3	197389	9	AL358791	AL358791 Human DNA
4	57	17.8	4424	9	HSPNP1	X79676 H.sapiens P
5	57	17.8	7546	9	HUMPTPB3	D21211 Human mRNA
6	57	17.8	8040	6	AR048218	AR048218 Sequence
7	57	17.8	8043	9	HSPTPL1	X80289 H.sapiens P
8	57	17.8	8062	9	HUMPTPB2	D21210 Human mRNA
9	57	17.8	8119	6	AR143153	AR143153 Sequence
10	57	17.8	8119	9	HUMPTPB1	D21209 Human mRNA
11	57	17.8	8287	9	HSU12128	U12128 Human prote
12	52.8	16.5	8411	4	BTU20807	U20807 Bos taurus
13	49.6	15.5	2433	9	BC010863	BC010863 Homo sapi
14	49.6	15.5	3956	9	HUMPTPSA	M83738 Human prote
15	47.2	14.8	4756	9	HUMPTPS	D37781 Human mRNA
16	47.2	14.8	5117	6	AR109903	AR109903 Sequence
17	47.2	14.8	5117	9	HSU10886	U10886 Human densi
18	46.6	14.6	2676	10	MMPTPPES	X87681 M.musculus
19	46.6	14.6	2983	10	MMPTPPTP	X63440 M.musculus
20	46	14.4	6314	10	RNU40790	U40790 Rattus norv
21	45.6	14.2	4315	9	HSLCAR	Y00638 Human mRNA
22	45.6	14.2	4597	9	HSLCA	Y00662 Human mRNA
23	45	14.1	2877	9	AF150732	AF150732 Homo sapi
24	44.2	13.8	2356	6	AX020263	AX020263 Sequence
25	44.2	13.8	2356	9	AF001847	AF001847 Homo sapi
26	44.2	13.8	3058	6	AX020261	AX020261 Sequence
27	44.2	13.8	3058	9	AF001846	AF001846 Homo sapi
28	44.2	13.8	3472	5	AF198450	AF198450 Xenopus 1
29	44.2	13.8	3580	6	AR149916	AR149916 Sequence
30	43.4	13.6	2321	10	RATBEM3	D45414 Rat mRNA fo
31	43.4	13.6	2674	10	AF332084	AF332084 Mus muscu
32	43.4	13.6	2683	10	AF332083	AF332083 Mus muscu
33	43.4	13.6	2952	10	RNICA105	X92563 R.norvegicu
34	43.4	13.6	3396	6	A37107	A37107 Sequence 1
35	43.4	13.6	3397	10	RNU40652	U40652 Rattus norv
36	43.4	13.6	3446	10	MMU11812	U11812 Mus musculu
37	43.4	13.6	3477	10	RATPDPTPLP	D38222 Rat mRNA fo
38	43.4	13.6	3561	6	A37109	A37109 Sequence 3
39	43.4	13.6	3566	10	MMPTP35A	X74438 M.musculus
40	43.2	13.5	3531	10	AF013490	AF013490 Mus muscu
41	43	13.4	1731	3	DDIPTPA	L15420 Dictyostell
42	43	13.4	1863	5	XELTYPHA	L33099 Xenopus lae
43	43	13.4	54594	9	AC073272	AC073272 Homo sapi
44	42.8	13.4	318	10	MMPTTYPHM	Z23060 M.musculus
45	42.4	13.2	1812	5	AB033587	AB033587 Potamo

#### ALIGNMENTS

##### RESULT 1

D64141  
LOCUS D64141  
DEFINITION Mus musculus mRNA for protein-tyrosine-phosphatase, complete cds.  
ACCESSION D64141  
VERSION D64141.1 GI:2665457  
KEYWORDS protein-tyrosine-phosphatase.  
SOURCE Mus musculus  
ORGANISM Mus musculus

##### REFERENCE

1 (bases 1 to 3090)  
Ohsugi, M.

##### AUTHORS

Direct Submission

##### JOURNAL

Submitted (15-SEP-1995) to the DDBJ/EMBL/GenBank databases. Miho Ohsugi, The University of Tokyo, Department of Oncology; 4-6-1 Shirokanedai, Minato-ku, Tokyo 108, Japan  
(E-mail: mhsugi@ims.u-tokyo.ac.jp, Tel: 03-5449-5305, Fax: 03-5449-5413)

##### REFERENCE 2

(sites)

AUTHORS Ohsugi, M., Kuramochi, S., Matsuda, S. and Yamanoto, T.  
TITLE Molecular cloning and characterization of a novel cytoplasmic protein-tyrosine phosphatase that is specifically expressed in spermatocytes  
JOURNAL J. Biol. Chem. 272 (52), 33092-33099 (1997)  
MEDLINE 98070510  
FEATURES Location/Qualifiers  
source 1..3090  
/organism="Mus musculus"  
/db\_xref="taxon:10090"  
/tissue\_type="testis"  
218..1498  
/EC\_number="3.1.3.48"  
/codon\_start=1  
/product="protein-tyrosine-phosphatase"  
/protein\_id="BAA23761.1"  
/db\_xref="GI:2665458"  
/translation="MSPSRKVRCKTGNDDEEGNSGNLNLRLNSLPSSQKMTPTKPI FGKMSNVKPSHLSFSDKYELVPEPLESDTETVWDSDRSURNRWNMSDETA GPSKTVSPVLGSSRLSKDTETSVSEKELTQLAQRPLIFNSARSAMRDLNLTQKK BELDIITRELELEQTPDPSNGNTLQNRKNRYDILPYSTRVPLGKNKDYINAS YRIVNHEEYFYATQGPPEPTIEDFWQVLENKCNVIAMITREIECGYKYSWP ISLKEPLEHFSVLETHVTOYFTVRVFOIVKSTGSKOCVKHLQFTKWPDPHGTPA SADFFTKYRVYRKSHITGELLVHCSAGVGRGVFCVDVWFSAEIKNYSFDIMNIVT OMKQRCGMIOHQEQYFCYEVLEVLQNLALY"  
BASE COUNT 904 a 603 c 647 g 936 t  
ORIGIN  
Query Match 87.5%; Score 280; DB 10; Length 3090;  
Best Local Similarity 92.2%; Pred. No. 8.2e-68;  
Matches 295; Conservative 0; Mismatches 25; Indels 0; Gaps 0;  
QY 1 GAAATAATTGTAATGTTATGCTATGATACACAGAGATAGAAAGTGGAGTTATCAAG 60  
Db 968 GAAATAATTGTAATGTTATGCTATGATACACAGAGATAGAAAGTGGAGTTATCAAG 1027  
QY 61 TGTTCAGATTACTGGCCGCTTCTCTCAAGAGAGCTTTGGAATCAACACTTTCATGTC 120  
Db 1028 TGTTCAGATTACTGGCCGCTTCTCTCAAGAGAGCTTTGGAATCAACACTTTCATGTC 1087  
QY 121 CTCTCGGAGAACTTCAGATAACCTCACTATTTTGTCCGAAATATGTCGATTTCAAAATGTCGAAG 180  
Db 1088 TTCTCGGAGAACTTCAGATAACCTCACTATTTTGTCCGAAATATGTCGATTTCAAAATGTCGAAG 1147  
QY 181 AAGTCCACAGAAAGAGTCACTCTGTAAACACTTGCAGTTCAATCAATGGCCAGACCAT 240  
Db 1148 AAGTCCACAGAAAGAGTCACTCTGTAAACACTTGCAGTTCAATCAATGGCCAGACCAT 1207  
QY 241 GGCACTCCTGCCTCAGTAGATTTTTCATCAATATGTCGTTATGTCGAGGAGGCCAC 300  
Db 1208 GGCACTCCTGCCTCAGTAGATTTTTCATCAATATGTCGTTATGTCGAGGAGGCCAC 1267  
QY 301 ATTACAGACCCCTCCTTGT 320  
Db 1268 ATTACAGACCCCTCCTTGT 1287  
RESULT 2  
HSM800374 2786 bp mRNA PRI 18-FEB-2000  
LOCUS Homo sapiens mRNA; cDNA DKFZp566K0524 (from clone DKFZp566K0524);  
DEFINITION partial cds.  
ACCESSION AL050040  
VERSION AL050040.1 GI:4884281  
KEYWORDS human.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 2786)  
AUTHORS Ansorge, W., Wirkner, U., Mewes, H.W., Gassenhuber, J. and Wiemann, S.  
TITLE Direct Submission

JOURNAL Submitted (15-MAY-1999) MIPS, Am Klopferspitz 18a, D-82152 Martinsried, GERMANY  
COMMENT Cloned from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@kfz-heidelberg.de; sequenced by EMBL (European Molecular Biology Laboratories, Heidelberg/Germany) within the cDNA sequencing consortium of the German Genome Project.  
This clone (DKFZp566K0524) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzd.de Further information about the clone and the sequencing project is available at http://www.mips.biochem.mpg.de/proj/cDNA/.  
FEATURES Location/Qualifiers  
source 1..2786  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="DKFZp566K0524"  
/clone\_lib="566 (synonym: hfk2). Vector pAMP1; host XL-2blue; sites NotI + SalI"  
/dev\_stage="fetal"  
/tissue\_type="kidney"  
1..1198  
/gene="DKFZp566K0524"  
/c1..1198  
/gene="DKFZp566K0524"  
/note="strong similarity to protein-tyrosine-phosphatases"  
/codon\_start=2  
/product="hypothetical protein"  
/protein\_id="CAB43248.1"  
/db\_xref="GI:4884282"  
/db\_xref="SPTREMBL:Q9Y406"  
/translation="GNWRGCGSGCTVGAAPQAVFENKVNSEKVLRLNPPHNDY EDVFEPSGSDPSMTARGLRDEWSEDEAAGSPQALSLSDTRKIVSEGL DLQAIPLIFNEHQTAIKDCLILEKTAAYDIMEQMALEKLNLPGEFYSGNPS NREKNRYDILPYDSTRVPLGKSDYINASYIRVNCGEYFYATQGPLLSTDDFW OMVLNNSNVIAMITREMEGGITKCVHYPISLKLPLKHPFVLENVQLLOYFIIR MQQVPEKSTGTSVSKLOFTKWPDPHGTPASADSFIKIRYARKSHLTGPMVHCSAG IGRGVFLCVDVWFSAEIKNYSFDIMNIVTOMKQRCGMIOHQEQYFCYEVLEVLRLNPPHNDY"  
polyA\_signal 2761..2766  
polyA\_site 2776  
BASE COUNT 841 a 494 c 550 g 901 t  
ORIGIN  
Query Match 75.5%; Score 241.6; DB 9; Length 2786;  
Best Local Similarity 84.7%; Pred. No. 4.3e-57;  
Matches 271; Conservative 0; Mismatches 49; Indels 0; Gaps 0;  
QY 1 GAAATAATTGTAATGTTATGCTATGATACACAGAGATAGAAAGTGGAGTTATCAAG 60  
Db 668 GAAATAATTGTAATGTTATGCTATGATACACAGAGATAGAAAGTGGAGTTATCAAG 727  
QY 61 TGTTCAGATTACTGGCCGCTTCTCTCAAGAGAGCTTTGGAATCAACACTTTCATGTC 120  
Db 728 TGTTCAGATTACTGGCCGCTTCTCTCAAGAGAGCTTTGGAATCAACACTTTCATGTC 787  
QY 121 CTCTCGGAGAACTTCAGATAACCTCACTATTTTGTCCGAAATATGTCGATTTCAAAATGTCGAAG 180  
Db 788 TTCTCGGAGAACTTCAGATAACCTCACTATTTTGTCCGAAATATGTCGATTTCAAAATGTCGAAG 847  
QY 181 AAGTCCACAGAAAGAGTCACTCTGTAAACACTTGCAGTTCAATCAATGGCCAGACCAT 240  
Db 848 AAGTCCACAGAAAGAGTCACTCTGTAAACACTTGCAGTTCAATCAATGGCCAGACCAT 907  
QY 241 GGCACTCCTGCCTCAGTAGATTTTTCATCAATATGTCGTTATGTCGAGGAGGCCAC 300  
Db 908 GGCACTCCTGCCTCAGTAGATTTTTCATCAATATGTCGTTATGTCGAGGAGGCCAC 967  
QY 301 ATTACAGACCCCTCCTTGT 320  
Db 968 ATTACAGACCCCTCCTTGT 987



/note="L1M4 repeat: matches 2651. .3102 of consensus"  
26879. .27248  
/note="L1M3 repeat: matches 5551. .5956 of consensus"  
27255. .27500  
/note="L1M1F repeat: matches 304. .568 of consensus"  
27667. .27736  
/note="L1M1F repeat: matches 3. .74 of consensus"  
27759. .27927  
/note="L1M3 repeat: matches 5992. .6162 of consensus"  
30024. .30310  
/note="L1M2 repeat: matches 1151. .1435 of consensus"  
31602. .31678  
/note="L1M1 repeat: matches 79. .155 of consensus"  
31627. .31927  
/note="match: STS: Em:HSJ10C10"  
31786. .31900  
/note="L1M1 repeat: matches 89. .203 of consensus"  
32440. .32738  
/note="L1M1 repeat: matches 1. .300 of consensus"  
32827. .32870  
/note="L1M1 repeat: matches 2 mer tt 75% conserved"  
34892. .35012  
/note="L1M2 repeat: matches 2584. .2710 of consensus"  
36225. .36549  
/note="L1M1 repeat: matches 21. .394 of consensus"  
36732. .36915  
/note="L1M1 repeat: matches 36. .234 of consensus"  
37945. .38060  
/note="L1M1 repeat: matches 48. .170 of consensus"  
38422. .38524  
/note="L1M1 repeat: matches 9. .116 of consensus"  
38641. .38785  
/note="L1M1 repeat: matches 34. .192 of consensus"  
39515. .39846  
/note="L1M3 repeat: matches 5845. .6182 of consensus"  
40561. .41030  
/note="L1M1 repeat: matches 72. .547 of consensus"  
41704. .41789  
/note="L1M1 repeat: matches 84. .166 of consensus"  
41923. .42236  
/note="L1M1 repeat: matches 5808. .6163 of consensus"  
45114. .45484  
/note="L1M1 repeat: matches 1. .365 of consensus"  
46495. .46591  
/note="L1M1 repeat: matches 43. .140 of consensus"  
47474. .47516  
/note="L1M1 repeat: matches 2658. .2700 of consensus"  
48316. .48775  
/note="L1M1 repeat: matches 6443. .6923 of consensus"  
49375. .49685  
/note="L1M1 repeat: matches 1. .299 of consensus"  
49686. .49720  
/note="L1M1 repeat: matches 445. .479 of consensus"  
49691. .49724  
/note="L1M1 repeat: matches 173. .206 of consensus"  
49725. .50193  
/note="L1M1 repeat: matches 19. .1523 of consensus"

Query Match 45.3%; Score 144.8; DB 9; Length 197389;  
Best Local Similarity 75.8%; Pred. No. 5.7e-30;  
Matches 179; Conservative 0; Mismatches 57; Indels 0; Gaps 0;  
QY 1 GAAATAATTGTAATGTTATGTTATGTAATACAGAGATAGAGTGGAGTTATCAAG 60  
|||||  
Db 161588 GAAATAATTGTAATGTTATGTTATGTAATACAGAGATAGAGTGGAGTTATCAAA 161529  
|||||  
QY 61 TGTTCAGTTACTGGCCGTTTCTCTCAAGAGAGCTTTGGAATTCACACACTTCATGTC 120  
|||||  
Db 161528 TGCTACCATTTACTGGCCATTTCTCTCAAGAGAGCTTTGGAATTCACACACTTCATGTC 161469  
|||||  
QY 121 CTTCGGAGAACTTTCAGATAACTCAGTATTTGTCATCCGAATATTTCAAAATTCGGAAG 180  
|||||  
Db 161468 TTCTCGGAGAACTTTCAGATAACTCAGTATTTGTCATCCGAATATTTCAAAATTCGGAAG 161409  
|||||

QY 181 AAGTCCACAGGAAGAGTCTCTGTAAACACTTGCAGTTTCATCAATGCCGAGA 236  
|||||  
Db 161408 AAGTCCGTAAGTTTTTAAACATCAGTATTAGATATATGTTTCCACCATAGCAGA 161353  
|||||  
RESULT 4  
HSPNP1 HSPNP1 4424 bp mRNA PRI 12-AUG-1996  
LOCUS H. sapiens PNP1 mRNA.  
X79676  
ACCESSION X79676  
VERSION X79676.1 GI:1486366  
KEYWORDS PNP1 gene; tyrosine phosphatase.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 4424)  
AUTHORS Wang, H.Y.  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 4424)  
AUTHORS Wang, H.Y.  
TITLE Direct Submission  
JOURNAL Submitted (08-JUN-1994) H.-Y. Wang, Max Planck Inst. fuer Biochemie,  
Dept of Molecular Biology, Am Klopferspitz 18A, 82152 Planegg, FRG  
FEATURES  
Location/Qualifiers  
1. .4424  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/tissue\_type="pancreas"  
<1. .3802  
/gene="PNP1"  
/product="tyrosine phosphatase"  
1. .4424  
/gene="PNP1"  
/product="tyrosine phosphatase"  
1. .4424  
/gene="PNP1"  
/codon\_start=2  
/product="tyrosine phosphatase"  
/protein\_id="CAA56124.1"  
/db\_xref="GI:1486367"  
/translation="DLSRSHCHVYLAHLENSFGPSGAREGLSSQDSRTESASLSQSQ  
VNGFFASHLGDTWQESQHSQSPSSVISKATEKETFDSQSKTKPGISDVTDYSDR  
GDSMDPEATYSSQDQHOTPKQESSSVNTSNKMNKTFPSSPPSGDIFEVLAKNDN  
GVILKNGASLKGQQEVIISALRGTAPEVLLCRPPGVLPEIDTALLTPQSPA  
QVLPNSSKDSQPSQVEQSTSDENEMSKKQCKSPKDSYSDSSGSGEDDLVTA  
PANISNTWSSALHOTLSNMVSOAOSHHEAPVKKIPFVPCFTILRKNKPEFEDSN  
PSPLPDMPAQSYQPSQESASSNDKYHIIHISEPTQENWTPLNKLNHLEDFE  
LEVLLITLIKSEKSGLFTVTKGNORIGYVHDVIOQPAKSDGPKVKKTKQVKKDYS  
DYTNMVAIVNLLKSKVRLVIRVLELPRPLPMLPDLITLTCNKEELGFSLCG  
GHDSLQVYVYISDINPRVAAIEGNQLLDVHYVNGVSTQGTMLLEEVNRLDMSLP  
LVLKATRNLDLPVPPSKRSAPKSTKNGSYSGVSGOPALTNDSTSTVAGEIN  
ELSPKKGCTVQIKGPNLTLPKESYIOEDDIYDQOAEVITQSLDDVDERSQNL  
NENNAAGYSGCGPCTLMKNGKLESEPTEDDCGSPLPFEYFEATKNGCEEYCEKVK  
SSYLOKPEKTKDDDEITWGNDELPIETRNHEDSDKHSF1TNDELAVLPVVKVLPS  
GSLTGANKLSVIRLVARSIGPSKELELQELKLDQCLIGTQKERNRNRYKNLP  
YDATRVPELGEKGYINAFIKIPVKEEYVYIACQGLPTTGVDFVQMTWEQKSTVIA  
MMTQVEGEKIKQRYWPNLTKTVMNRLALVLMQOLKGVVRAWMTLEDIQFRE  
VRHSHLNTAMPDHTPSQDDLLTFISYMRH1HRSQPTLTHCSAGIGRSGLICID  
VVLGLISQDLDFDISDLVRCMLRQHGVMQVTEOYIFCYQVILYVLTQLQAEQKQ  
PQLLK"

BASE COUNT 1452 a 909 c 922 g 1141 t  
ORIGIN  
Query Match 17.8%; Score 57; DB 9; Length 4424;

	Best Local Similarity	50.8%;	Pred. No.	1.5e-05;	
	Matches	163;	Conservative	0;	Mismatches 155; Indels 3; Gaps 1;
QY	1	GAAATAATGTAATTGTTATTCGTATCATACAGAGAGATAGAAGTGGAGTTATCAAG	60		
Db	3242	GAGCAAAAATCCACAGTGATAGCCTGATGACTCAGAGTAGAGAGAGAAAAAATCAA	3301		
QY	61	TGTTGCAGTTACTGGCGCGTTTCCT--GAAGGAGGCTTTGGAAATTCACAACATTTG	117		
Db	3302	TGCCAGCGCTATTGGCGCAACATCCTAGGCAAAAACAATGGTCAGCAACAGACTT	3361		
QY	118	GTCTTCTGGAGAACCTTCAGATAACCTCAGTATTTTTGTGCATCCGAATATTTCA	177		
Db	3362	CTGCTCTCTGTGAGAAATGACAGCAGCTGAAGGCGCTTTGTGGTGAGGCGCAAT	3421		
QY	178	AAGAAGTCCACAGAGAAAGAGTCACTCTGTAAACACATTTGCAGTTTCATCAAA	237		
Db	3422	GATATTCAGACCAGAGAGGTGGCGCATATTTCTCATCTGATTTCACTGCTGCC	3481		
QY	238	CATGGCACTCTCTGCTCAGTAGATTTTTTCATCAAAATATGTCGCTTTATGTG	297		
Db	3482	CATGATACACCTTCTCAACAGAGATGATCTGCTTACTTTTATCTCTACATGAC	3541		
QY	298	CACATTACAGACCCCCTT	318		
Db	3542	CACAGATCAGGCCCAATCAT	3562		
RESULT	5				
HUMPTPB3					
LOCUS	HUMPTPB3	7546 bp	mRNA	PR1	05-FEB-1999
DEFINITION					Human mRNA for protein tyrosine phosphatase (PTP-BAS, type 3), complete cds.
ACCESSION	D21211				
VERSION	D21211.1	GI:452193			
KEYWORDS	PTP-BAS; protein tyrosine phosphatase.				
SOURCE	Homo sapiens (library: Lambda zap II) basophilic leukemia cells cell-line K0812E cDNA to mRNA, clones BAS-[1-4].				
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE	1	(bases 1 to 7546)			
AUTHORS	Maekawa,K.				
TITLE	Direct Submission				
JOURNAL	Submitted (15-OCT-1993) to the DDBJ/EMBL/GenBank databases. Kazuhiko Maekawa, Shionogi Institute for Medical Science; 2-5-1 Mishima, Settsu, Osaka 566, Japan (Tel:06-382-2612, Fax:06-382-2598).				
REFERENCE	2	(bases 1 to 7546)			
AUTHORS	Maekawa,K., Imagawa,N., Nagamatsu,M. and Harada,S.				
TITLE	Molecular cloning of a novel protein-tyrosine phosphatase containing a membrane-binding domain and GLGF repeats				
JOURNAL	FEBS Lett. 337 (2), 200-206 (1994)				
MEDLINE	94116679				
COMMENT	Submitted (15-Oct-1993) to DDBJ by: Kazuhiko Maekawa Shionogi Institute for Medical Science 2-5-1 Mishima, Settsu Osaka 566 Japan Phone: 06-382-2612 Fax: 06-382-2598.				
FEATURES					
source	Location/Qualifiers				
	1..7546				
	/organism="Homo sapiens"				
	/db_xref="taxon:9606"				
	/cell_line="K0812E"				
	/cell_type="basophilic leukemia cells"				
	/clone_lib="Lambda zap II"				
	64..6948				
	/codon_start=1				
CDS	/product="protein tyrosine phosphatase (PTP-BAS, type 3)" /protein_id="BAA04752.1"				





```

DB 7258 CACAGATCAGGCCCAATCAT 7278
||||| ||||| || || ||
RESULT 10
HUMTPB1
LOCUS
DEFINITION
HUMTPPB1 8119 bp mRNA PRI 05-FEB-1999
Human mRNA for protein tyrosine phosphatase (PTP-BAS, type 1),
complete cds.
ACCESSION
D21209.1 GI:452189
PTP-BAS; protein tyrosine phosphatase.
KEYWORDS
Homo sapiens (library: Lambda Zap II) basophilic leukemia cells
SOURCE
cell-line KU812E cDNA to mRNA, clones BAS-[1-4].
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 8119)
Maekawa,K.
Direct Submission
Submitted (15-OCT-1993) to the DDBJ/EMBL/GenBank databases.
Kazuhiko Maekawa, Shionogi Institute for Medical Science; 2-5-1
Mishima, Settsu, Osaka 566, Japan (Tel:06-382-2612,
Fax:06-382-2598)
2 (bases 1 to 8119)
Maekawa,K., Inagawa,N., Nagamatsu,M. and Harada,S.
Molecular cloning of a novel protein-tyrosine phosphatase
containing a membrane-binding domain and GLGF repeats
FEBS Lett. 337 (2), 200-206 (1994)
94116679
Submitted (15-OCT-1993) to DDBJ by:
Kazuhiko Maekawa
Shionogi Institute for Medical Science
2-5-1 Mishima, Settsu
Osaka 566
Japan.
Phone:
06-382-2612
Fax:
06-382-2598.
FEATURES
Location/Qualifiers
1..8119
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="KU812E"
/cell_type="basophilic leukemia cells"
/clone_lib="Lambda zap II"
64..7521
/codon_start=1
/product="protein tyrosine phosphatase (PTP-BAS, type 1)"
/protein_id="BAA04750.1"
/db_xref="GI:452190"
/translation="MHVSLAEALVVRGGLPQEEIHWLVNQASIEQLQELFRKVSADP
AALGFTISPSWLLLPSSGVSTDENISNODLRAFTAPELVNQSLTSLSDVEKHIIY
NCAGPSTYVKHLVKVQSPQIKGLDHLSILSCGEQDIYARVSVRTLQACSAHINS
SLGPMSTYKGLVGLVQSPQIKGLDHLSILSCGEQDIYARVSVRTLQACSAHINS
VLDIQKPLSHQTFLLNGLSKSGMGLSDIQENYFKOILSDNSGRESENTFTSPYQ
FTTPRGEKAPIPGIDVLISKKKIWASSMDLLCTADRFSSGATATYRRCHPEAVTVTS
FPTKREKAPISDGLSIDLIFGQKMDPIYHTRLEPTSSAISALDRIRQKKLQVDR
EAMNVEEPRYKTYHGVSTSESPSIISSSEDFQRVRSEAKRFRSSGGLQVDR
ETLSQSQSPQRKYETPEGNLQNEIMLKRQEEELMQLQAKMARQLRSLYPGDT
IKASMLDITRDLPLETALMTAKLRNFPFGFVKMTIEPIISDLPRSLTKKKGK
NEDNRKKNVIMLLNGORLETCDTKTICKDQFDMVVAHGLVPEHHLFALATLKDNFY
FYDPPDLKLFKVAPECKHEPKKTKATVNTFFRRIKPFMDVDSVLQHTLTCHOYVLQ
LRKDILIEERHCDDETSLLSALAQAGDYQDPEVHGVSYFRMEHYLPARVMEKLDL
SYIKPELKLNTYVGASEKTELEELFKYCRLTGEYGVHFRVPEPKQTDGILLVC
SKGLVFEYVHNGVRTLVLRFPPWRETKKISFKKKITLQNTSDGIKHGFTQDNKIQCY
LHLGYSQHKFOLQMRARQSNODADIERASFRSLNLAQESVRGNMGRAISTGLAS
STLNKLAVPLRQVQAPILKRLSCSELISYQPLQNSKKNDAESAGVTKLNKSKSVASL
LSQASLYPRKKNVIMNPPROTVAELVCKPSSQNSKDAESAGVTKLNKSKSVASL
REITLVNKKDKAGLUGLQIDGKRMKRDLDLGISSVAPGGADLGGCLKPGDRLLIS
NNSPERRKIESDSSSTEDFGQAVLGNTMHHSSNSSSQVFLKENDVLRHNSLVSPSE
VNSVLGVGSHAAIEILQNAPEEDMTVLISQPEKISYKVPSTPVHILNEMKNYMKKS
YMODASIDSSKDKHWSRGTLRHISSEFCSQGLREGSLSDQSRTESASQSQVFN
GFASFHGQTDWOESOHGSPSPISKATEKFTFTDQNSQTKKPGISDVDTQYSDRGD

```





conflict  
 conflict 4368..4382  
 /gene="PTPLP"  
 /notes="nucleotides 4368..4382 not present in sequence of PTP-BAS1 gene, GenBank Accession Number D21209"

3'UTR  
 polyA\_site  
 7691..8287

BASE COUNT 2643 a 1694 c 1780 g 2170 t  
 ORIGIN

Query Match 17.8%; Score 57; DB 9; Length 8287;  
 Best Local Similarity 50.8%; Pred. No. 1.6e-05;  
 Matches 163; Conservative 0; Mismatches 155; Indels 3; Gaps 1;

```

QY      1   GAAATAATTTGTAATGTATTTCCTATCAACAGAGAGATAGAAGTGCAGTTATTC AAG 60
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db       7127 GAGCAAAAATCCACAGTGATGCGCATGCTCAAGAAGTAGAGAGAAAAATC AAA 7186

QY      61   TGTTCAGTACTAGGCCGGCTTTCTCT---GAAGAGAGCCTTTGGAAATTC AAACACTTTTCAT 117
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db       7187 TGCCAGCGCTATTGGCCCAACATCTTAGGCAAAACAACAATGGTCAGCAACAGACTTCGA 7246

QY      118   GTCTTCTCGAGAACCTTTCAGATAAECTCACATATTTTCTCATCCGANTATTTC AAANTGTG 177
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db       7247 CTGCTCTTGTGAGATGCAGAGCTGAAGGCGCTTTGTGTGAGGCGCAATGACCCTTTGAA 7306

QY      178   AGAAGTCCACAGGAAGAAGTCACTCTGTAAAAACACTTGCAGTTTCATCAAATGCCAGAC 237
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db       7307 GATATTGAGACAGAGAGGTGCGCATATTTCTCATCTGAATTTCTACTGCCTGGCCAGAC 7366

QY      238   CATGGCACTCCTGCCCTCAGTAGATTTTTTTCATCAATATGTCGGTTATGTGAGAGAGAC 297
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db       7367 CATGATACACTCTCTCAACAGATGATCTGCTTACTTTTATCTCTACATGAGACACATC 7426

QY      298   CACATTACAGAGCCCCTCCTT 318
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db       7427 CACAGATCAGGCCCAATCATTT 7447
  
```

RESULT 12

LOCUS	BTU20807	8411 bp	mRNA	MAM	29-JUL-1995
DEFINITION	Bos taurus protein tyrosine phosphatase BA14 mRNA, complete cds.				
ACCESSION	U20807				
VERSION	U20807.1	GI:915209			
KEYWORDS	cow.				
SOURCE	Bos taurus				
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.				
REFERENCE	1 (bases 1 to 8411)				
AUTHORS	Vega,Q.C., Walton,K.M. and Dixon,J.E.				
TITLE	Tyrosine Phosphatases and the Cytoskeleton				
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 8411)				
AUTHORS	Walton,K.M.				
TITLE	Direct Submission				
JOURNAL	Submitted (09-FEB-1995) Kevin M. Walton, Biological Chemistry, University of Michigan, 1301 E. Catherine, Ann Arbor, MI 92093-0606, USA				
FEATURES	Location/Qualifiers				
source	1..8411				



```
BASE COUNT      953 a   1000 c   992 g   1011 t
ORIGIN
Query Match      15.5%; Score 49.6; DB 9; Length 3956;
Best Local Similarity 48.3%; Pred. No. 0.0018;
Matches 139; Conservative 0; Mismatches 149; Indels 0; Gaps 0;
QY 16 GTTATTCCTATGATACCAGAGAGATAGAAAGGTGGAGTTATCAAGTGTTCAGTTACTGG 75
Db 1697 GTGATTTCATGACACCCCGCTTTCAGGAAGGGCGGAGGAGAAAGTGGCCAGTACTGG 1756
QY 76 CCGCTTCTCGAGGAGCCCTTGGAAATTCACACATTTTCATGCTCCTCTCGGAACTTT 135
Db 1757 CTTTAGAAAAGACTCTCGGATCCGATTTGGCTTCTCACAGTACCAATCTAGGCGTG 1816
QY 136 CAGATACTCAGTATTTGTGATCCGAAATATTTCAAAATTTGGAAGAAGTCCACAGGAAAG 195
Db 1817 GAGAACATGAATCATATAAGAAAACACGCTAGAAATTCACACACAGAGGACGCGAG 1876
QY 196 AGTCACCTCTGTAACACACTTGCAGTTCATCAAAATGGCCAGACCATGGCACTCTCTGCCTCA 255
Db 1877 AAAGCGCAGGTGACCCACTTCCAGTTCCTGAGCTGGCCAGACTATGTTGCTCCTCA 1936
QY 256 GTAGATTTTTCATCAATATATGCCGTTATGTGAGGAGAGGCCACATT 303
Db 1937 GCAGCTTCCCTCATGACTTCTTGAGAGTGGTGCAGAAACAGCAGACT 1984

RESULT 15
LOCUS HUMTPPS 4756 bp mRNA PRI 01-FEB-2000
DEFINITION Human mRNA for protein-tyrosine phosphatase HPTPeta, complete cds.
ACCESSION D37781
VERSION D37781.1 GI:633072
KEYWORDS protein-tyrosine phosphatase; HPTPeta.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 4756)
AUTHORS Honda,H., Inazawa,J., Nishida,J., Yazaki,Y. and Hirai,H.
TITLE Molecular cloning, characterization, and chromosomal localization
of a novel protein-tyrosine phosphatase, HPTP eta
JOURNAL Blood 84 (12), 4186-4194 (1994)
MEDLINE 95086212
REFERENCE 2 (bases 1 to 4756)
AUTHORS Honda,H.
TITLE Direct Submission
JOURNAL Submitted (23-JUL-1994) to the DDBJ/EMBL/GenBank databases.. Hiroaki
Honda, Jichi Medical School, Department of Molecular Biology;
3311-1 Yakushiji, Minamikawachi-machi, Kawachi-gun, Tochigi 329-04,
Japan (Tel:+81-285-44-2111(ex.3482), Fax:+81-285-44-8675)
FEATURES
source 1..4756
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..343
345..4358
/codon_start=1
/product="protein-tyrosine phosphatase"
/protein_id="BAA07035.1"
/db_xref="GI:633073"
/translation="MKPAAREARLPPRPLRWALPLLLLRLQLILCAGTTPSPIP
DPSVATGENTQISIAESFHKQNGTGPVETNTSEDCSSGANDSLRTPQ
SNGTDGASOKTSPSPGPFDKAVSISPTNVLTKNSDNTAASEYKYVWKHME
KITIVVHPWCNITGLRPATSYVESITPGIGNETGCDPRVIVTEPIVPSDLRVALT
GYRKAALSNNGTASCRLLESIGSHEELTQDSRLQVNSGLAPGVQYINPILQ
SNKTGRLPTGGDLASNTSRAGSPAPVHDSLVGVPSSGQQSRDTEVLVVG
LEPGTRYNATVYQAANGQGPQAFETFTNAIQVDFVAVNISATSLIWKVSDNE
SSSNPTKTHVAGETDSSNLNISEPRAVTPGSRSTFYNTVTCVILGIEGTPGLQV
HTPPVPVDFRTVTVSTTEIGLAWSHDAESFQMHTIQEGAGNSRVEITNQSIIGG
LFPFGTKYCEIIVKPGNPTGASRTVCNRTVPVSAVEDIHVVVVTITTEMWLDKNSPDGA
```

```
SEYVYHLVIESKHGSHNTSTYDKAITLQGLIPGTLNITISPEVDHVMGDPNSTAQYT
RPSNVSNIDVSTNTAATLSQNFDDASPTYSYCLLIERAGNSNATQVVTDIGITDA
TPTLIPGSGSYTVEIFAQVCDGKISLEPGRKSFCTDPASMPDCEVVPKEPALVKW
TCPPGANAGFELEVSSGANNATHLESCSENTEYRTEVTYNLFSTSYNISITTVSC
GKMAAPTRNTCTTGTIDPPPDGSPNITSVSHNSVKVKTSGFEASHGPIKAYAVILTT
GEAGHPSDVLYKYTYDDFKKASDTYVYLIRTEEGRSQSLSEVLYEIDVGNESTT
LGLQWEAGTSGLLPACVAGFTIFHPQNKGLIDGAEYSVSFSRYSKSLIRVENFEAYF
ICGAVEGCFGALVIVTVGGFIFWRKKRDKAKNEVSFSQIKPKKSLIRVENFEAYF
KKQADNSNGCFAEYEDLKLVGISQPYAAELAKENGRKNRYNNVLPYDLSRVKLSVQT
KKQDDYINANYMPGYHKKDFIATQGPLNLTLDKWRVMYKKNVYALIMLTKEVEQGR
TKCEEVWSPSKOQDYGDIIVANTSEIVLPWTIRDTFVNKIOTSESHPLRPHFTSKP
DHGVPDPTDLLINFRYLVRDVKQSPESPILVHCSAGVGRGTGTFTAIDRLIYQINE
NVDVYGVIVDILRHRPLMWQTEQYVFLNQCVLDIVRSQKSKVLDLIYONTAMTYI
ENLAPVITTFKNGYIA"
sig_peptide 399..458
transit_peptide 3255..3329
3'UTR 4355..4756
BASE COUNT 1241 a 1207 c 1220 g 1088 t
ORIGIN
Query Match      14.8%; Score 47.2; DB 9; Length 4756;
Best Local Similarity 51.1%; Pred. No. 0.0085;
Matches 140; Conservative 0; Mismatches 128; Indels 6; Gaps 1;
QY 1 GAAATAAATTTGTAATGTTATGCTATGATAACAGAGAGATAGAGGTGGAGTTATCAAG 60
Db 3729 GAGAAAAATGTATATGCCATCATTTATGTTGACTAAATGTGTGAACAGGGAAGAACAAA 3788
QY 61 TGTGTCAGTTACTGGCCCGTTTCTCTGAAGGAGCGCTTTGGNAATTCAAACACTTTTCATGTC 120
Db 3789 TGTGAGGAGTATTTGGCCCC-----TCCAAGCAGGCTCAGGACTATGGAGACATAAATGTTG 3842
QY 121 CTCTCTGAGAACTTTCAGATAAATCACTAGTATTTTCTCATCCGAATATTTCAAAATTTGTCGAAG 180
Db 3843 GCAATGACATCAGAAATTTGTTCTCCGAATGGACCATCAGAGATTTTCAGTGAATAAT 3902
QY 181 AGTCCACAGGAAGAGTCACTCTGTAAAACACTTGCAGTTGCATCAATGCCAGACCAT 240
Db 3903 ATCCAGACAAAGTGAAGTCACTCCCTCTGAGACAGTTCACCTTTCACCTCTGCGCCAGACCAC 3962
QY 241 GGCACCTCTCCCTCAGTAGATTTTTTTCATCAAAAT 274
Db 3963 GGTGTTCCCGACACACCACTGACCTGCTCATCACT 3996
Search completed: March 30, 2002, 05:16:13
Job time: 3331 sec
```

---

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2002, 04:22:52 ; Search time 171.25 Seconds  
(without alignments)  
1602.009 Million cell updates/sec

Title: US-09-095-478A-4  
Perfect score: 320  
Sequence: 1 GAAATAATGTAATGTTAT.....ATTACAGGACCCCTCTTGT 320

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 10%  
Listing first 45 summaries

Database : N\_Geneseq\_1101.\*  
1: /SIDS2/gcgdata/geneseq/geneseq/NA1980.DAT:\*  
2: /SIDS2/gcgdata/geneseq/geneseq/NA1981.DAT:\*  
3: /SIDS2/gcgdata/geneseq/geneseq/NA1982.DAT:\*  
4: /SIDS2/gcgdata/geneseq/geneseq/NA1983.DAT:\*  
5: /SIDS2/gcgdata/geneseq/geneseq/NA1984.DAT:\*  
6: /SIDS2/gcgdata/geneseq/geneseq/NA1985.DAT:\*  
7: /SIDS2/gcgdata/geneseq/geneseq/NA1986.DAT:\*  
8: /SIDS2/gcgdata/geneseq/geneseq/NA1987.DAT:\*  
9: /SIDS2/gcgdata/geneseq/geneseq/NA1988.DAT:\*  
10: /SIDS2/gcgdata/geneseq/geneseq/NA1989.DAT:\*  
11: /SIDS2/gcgdata/geneseq/geneseq/NA1990.DAT:\*  
12: /SIDS2/gcgdata/geneseq/geneseq/NA1991.DAT:\*  
13: /SIDS2/gcgdata/geneseq/geneseq/NA1992.DAT:\*  
14: /SIDS2/gcgdata/geneseq/geneseq/NA1993.DAT:\*  
15: /SIDS2/gcgdata/geneseq/geneseq/NA1994.DAT:\*  
16: /SIDS2/gcgdata/geneseq/geneseq/NA1995.DAT:\*  
17: /SIDS2/gcgdata/geneseq/geneseq/NA1996.DAT:\*  
18: /SIDS2/gcgdata/geneseq/geneseq/NA1997.DAT:\*  
19: /SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT:\*  
20: /SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT:\*  
21: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:\*  
22: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	320	100.0	320	20 AAV81747	Rat PTP10 encoding
2	280	87.5	1692	20 AAV81746	Mouse PTP05 isofor
3	280	87.5	1785	20 AAV81744	Mouse PTP05 encodi
4	280	87.5	1896	20 AAV81745	Mouse PTP05 isofor
5	57	17.8	8043	16 AAQ85924	Human protein tyro
6	57	17.8	8119	21 AAC61840	DNA encoding a hum
7	47.2	14.8	5117	16 AAT06027	huDEP-1 cDNA. Hom
8	45	14.1	936	22 AAF58252	Oligonucleotide D1
9	45	14.1	936	22 AAF58254	Oligonucleotide D1
10	45	14.1	936	22 AAF58257	Oligonucleotide D1
11	45	14.1	936	22 AAF58259	Oligonucleotide D2

C 12	45	14.1	936	22	AAF58262	Oligonucleotide D2
C 13	45	14.1	938	22	AAF58255	Oligonucleotide D1
14	44.6	13.9	4798	22	AAI58418	Human polynucleoti
15	44.4	13.9	936	22	AAF58252	Oligonucleotide D1
16	44.4	13.9	936	22	AAF58254	Oligonucleotide D1
17	44.4	13.9	936	22	AAF58257	Oligonucleotide D1
18	44.4	13.9	936	22	AAF58259	Oligonucleotide D2
19	44.4	13.9	936	22	AAF58262	Oligonucleotide D2
20	44.4	13.9	938	22	AAF58255	Oligonucleotide D1
21	44.2	13.8	2356	20	AAI90696	Human Cytoplasmic
22	44.2	13.8	3058	20	AAI90695	Human Cytoplasmic
23	44.2	13.8	3580	20	AAV81742	Human PTP04 encodi
24	43.4	13.6	3396	15	AAQ55515	PPase PTP35 genom
25	43.4	13.6	3561	17	AAQ55970	PPase PTP35 DNA s
26	38.8	12.1	285	15	AAT35309	Rat M2PTP gene cDN
27	36.6	11.4	1191	20	AAI86091	Nucleotide sequenc
28	36.6	11.4	1210	18	AAI95246	Human IA-2/ICA512
29	36.6	11.4	1344	20	AAI86092	Nucleotide sequenc
30	36.6	11.4	1413	20	AAI23241	Human pancreatic i
31	36.6	11.4	1413	20	AAI209922	Human islet cell a
32	36.6	11.4	1413	20	AAI63560	Islet cell antibod
33	36.6	11.4	1413	21	AAI46459	pancreatic islet c
34	36.6	11.4	3311	20	AAI23242	Human pancreatic i
35	36.6	11.4	3311	20	AAI209923	Human islet cell a
36	36.6	11.4	3311	20	AAI63561	Islet cell antibod
37	36.6	11.4	3311	21	AAI46460	Pancreatic islet c
38	36.6	11.4	3613	20	AAI87627	Human islet cell c
39	36.6	11.4	3984	18	AAI58627	Protein tyrosine p
40	36.6	11.4	3984	20	AAI78463	Human PTPH1 cDNA
41	36.4	11.4	244	22	AAI58238	Oligonucleotide D1
42	34.8	10.9	244	22	AAI58238	Oligonucleotide D1
43	34.4	10.8	626	21	AAI44264	Porcine IFN-gamma
44	34.4	10.8	842	19	AAI08186	Porcine interferon
45	34.4	10.8	842	20	AAI63230	Porcine interferon

## ALIGNMENTS

RESULT 1  
AAV81747  
ID AAV81747 standard; cDNA; 320 BP.  
XX AAV81747;  
AC AAV81747;  
DT 10-MAR-1999 (first entry)  
XX Rat PTP10 encoding cDNA.  
DE Rat PTP10 encoding cDNA.  
XX PTP04; PTP05; PTP10; SAD; ALP; ALK-7; protein tyrosine phosphatase;  
KW type I receptor serine/threonine kinase; cancer; leukaemia; lymphoma;  
KW neurodegenerative disease; neuronal survival; Alzheimer's disease;  
KW Parkinson's disease; Huntington's disease; ss.  
XX Rattus sp.  
XX WO9849317-A2.  
PD PD  
XX 05-NOV-1998.  
XX 27-APR-1998; 98WO-US08439.  
XX 23-OCT-1997; 97US-0063595.  
PR 28-APR-1997; 97US-0044428.  
PR 20-MAY-1997; 97US-0047222.  
PR 11-JUN-1997; 97US-0049477.  
PR 11-JUN-1997; 97US-0049756.  
PR 18-JUN-1997; 97US-0049914.  
XX (SUGEN-) SUGEN INC.  
PA App H, Clary D, Courtneidge SA, Hui TH, Jallal B;  
PI Markby D, Onrust S, Peles E, Plowman GD;

```

XX WPI: 1999-009434/01.
DR P-PSDB; AAW89252.
XX
XX New nucleic acid encoding specific protein tyrosine phosphatases -
PT useful for identifying specific modulators for treatment and
PT prevention of cancer and neurodegenerative disease
XX
XX Claim 2; Page 149; 193pp; English.
PS
XX The present invention describes isolated, enriched or purified nucleic
CC acids encoding PTP04, SAD, PTP05, PTP10, ALP and ALK-7 proteins. The
CC present sequence encodes rat PTP10. The above proteins, other than
CC ALK-7, are protein tyrosine phosphatases (PTPs) and are used to identify
CC substances that modulate their activity (i.e. agonists and antagonists,
CC including NBP) in vivo or in vitro. These substances are used to treat
CC or prevent diseases associated with abnormal signal transduction
CC pathways that involve the proteins, particularly cancer (e.g. leukaemia
CC and lymphoma), while modulators of ALK-7 (which is a type I receptor
CC serine/threonine kinase) are used to promote neuronal survival,
CC particularly for treating Alzheimer's, Parkinson's or Huntington's
CC diseases. Nucleic acid fragments of the polynucleotides encoding the
CC proteins can be used as probes to identify and clone related sequences;
CC to detect protein-encoded RNA; to generate transgenic animals and in
CC gene therapy (optionally after mutation). Ab are used to determine the
CC proteins.
XX
XX Sequence 320 BP; 95 A; 64 C; 65 G; 96 T; 0 other;

Query Match 100.0%; Score 320; DB 20; Length 320;
Best Local Similarity 100.0%; Pred No. 6.4e-89;
Matches 320; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAAATAATTTGTAATGTTATGCTATGATAACCCAGAGAGATAGAGGTGGAGTTATCAAG 60
DB 1 gaaataatgttaatgttattgctatgataaccagagagatagaaggtggagttatcaag 60
QY 61 TGTGTGAGTTACTGGCCCGTTTCTCTGAAGAGCCCTTGGAAATCAACACTTTCATGTC 120
DB 61 tgttgtagttactggcccggttctctgaagagcccttggaaatcaaacactttcatgtc 120
QY 121 CTCTGTGAGAACTTTCAGATAACTCAGTATTTGTCTATCCGAATATTTCAAAATGTGAAG 180
DB 121 ctctgtgagaacttttcagataactcagatatttgtctatccgaatatttcaaaatgtgaag 180
QY 181 AAGTCCACAGAAAGAGTCTACTCTGTAACACTTGCAGTTCATCAATGCCAGACCAT 240
DB 181 aagtccacagaaagagtcactctgttaaacacttgcagttcatcaaatgccagaccat 240
QY 241 GGCACCTCTCCCTCAGTATTTTTCATCAATATGTCCTCGTTATGTGAGGAGGCCAC 300
DB 241 ggcactctctccctcagtagatttttcatacaaatatgtccgttatgtgaggaagccac 300
QY 301 ATTACAGGACCCCTCTCTGT 320
DB 301 attacaggaccctctctgt 320

RESULT 2
ID AAV81746
XX AAV81746 standard; cDNA; 1692 BP.
AC
XX AAV81746;
XX
XX 10-MAR-1999 (first entry)
DE Mouse PTP05 isoform #2 encoding cDNA.
XX
XX PTP04; PTP05; SAD; ALP; ALK-7; protein tyrosine phosphatase;
KW type I receptor serine/threonine kinase; cancer; leukaemia; lymphoma;
KW neurodegenerative disease; neuronal survival; Alzheimer's disease;
KW Parkinson's disease; Huntington's disease; ss.

```

```

XX Mus sp.
XX OS
XX FH Key Location/Qualifiers
XX CDS 199..1413
XX FT /*tag= a
XX FT /note= "no stop codon given"
XX
XX WO9849317-A2.
XX
XX 05-NOV-1998.
XX
XX 27-APR-1998; 98WO-US08439.
XX
XX 23-OCT-1997; 97US-0063595.
XX 28-APR-1997; 97US-0044428.
XX 20-MAY-1997; 97US-0047222.
XX 11-JUN-1997; 97US-0049477.
XX 11-JUN-1997; 97US-0049756.
XX 18-JUN-1997; 97US-0049914.
XX
XX (SUGE-) SUGEN INC.
XX
XX App H, Clary D, Courtneidge SA, Hui TH, Jallal B,
XX Markby D, Onrust S, Peles E, Plowman GD;
XX
XX WPI: 1999-009434/01.
XX P-PSDB; AAW89251.
XX
XX New nucleic acid encoding specific protein tyrosine phosphatases -
PT useful for identifying specific modulators for treatment and
PT prevention of cancer and neurodegenerative disease
XX
XX Claim 2; Page 148-149; 193pp; English.
XX
XX The present invention describes isolated, enriched or purified nucleic
CC acids encoding PTP04, SAD, PTP05, PTP10, ALP and ALK-7 proteins. The
CC present sequence encodes mouse PTP05. The above proteins, other than
CC ALK-7, are protein tyrosine phosphatases (PTPs) and are used to identify
CC substances that modulate their activity (i.e. agonists and antagonists,
CC including NBP) in vivo or in vitro. These substances are used to treat
CC or prevent diseases associated with abnormal signal transduction
CC pathways that involve the proteins, particularly cancer (e.g. leukaemia
CC and lymphoma), while modulators of ALK-7 (which is a type I receptor
CC serine/threonine kinase) are used to promote neuronal survival,
CC particularly for treating Alzheimer's, Parkinson's or Huntington's
CC diseases. Nucleic acid fragments of the polynucleotides encoding the
CC proteins can be used as probes to identify and clone related sequences;
CC to detect protein-encoded RNA; to generate transgenic animals and in
CC gene therapy (optionally after mutation). Ab are used to determine the
CC proteins.
XX
XX Sequence 1692 BP; 507 A; 342 C; 386 G; 457 T; 0 other;

Query Match 87.5%; Score 280; DB 20; Length 1692;
Best Local Similarity 92.2%; Pred. NO. 2.4e-76;
Matches 295; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

QY 1 GAAATAATTTGTAATGTTATGCTATGATAACCCAGAGATAGAGGTGGAGTTATCAAG 60
DB 949 gaaataatgttaatgttattgctatgataaccagagagatagaaggtggagttatcaag 1008
QY 61 TGTGTGAGTTACTGGCCCGTTTCTCTGAAGAGCCCTTGGAAATCAACACTTTCATGTC 120
DB 1009 tgttgtagttactggcccggttctctgaagagcccttggaaatcaaacactttagtc 1068
QY 121 CTCTGTGAGAACTTTCAGATAACTCAGTATTTGTCTATCCGAATATTTCAAAATGTGAAG 180
DB 1069 tctctgtgagaacttttcagataactcagatatttgtctatccgaatatttcaaaatgtgaag 1128
QY 181 AAGTCCACAGAAAGAGTCTACTCTGTAACACTTGCAGTTCATCAATGCCAGACCAT 240
DB 181 aagtccacagaaagagtcactctgttaaacacttgcagttcatcaaatgccagaccat 240

```



Db 1129 aagtcacagaaagagccaatgtgtataaacacattgcagttccaccagagccagaccat 1188  
 Qy 241 GGCACTCTGCTCAGTAGATTTTTCATCAATATATGTCCTTATGTGAGGAAGAGCCAC 300  
 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 1189 ggcactctgctcagcagatgtttttccataaaatgctccgttatgtgaggaagagccac 1248

Qy 301 ATTACAGAGCCCTCCTTGT 320  
 |||||||||||||||||||||||  
 Db 1249 attacaggagccctccttgt 1268

## RESULT 3

AAV81744  
 ID AAV81744 standard; cDNA; 1785 BP.

AC AAV81744;

DT 10-MAR-1999 (first entry)

DE Mouse PTP05 encoding cDNA.

XX PTP04; PTP05; PTP10; SAD; ALP; ALK-7; protein tyrosine phosphatase;  
 KW type I receptor serine/threonine kinase; cancer; leukaemia; lymphoma;  
 KW neurodegenerative disease; neuronal survival; Alzheimer's disease;  
 KW Parkinson's disease; Huntington's disease; ss.

OS Mus sp.

XX Key Location/Qualifiers  
 FH 199..1479  
 FT CDS /\*tag= a  
 XX

PN W09849317-A2.

XX 05-NOV-1998.

XX 27-APR-1998; 98WO-US08439.

XX 23-OCT-1997; 97US-0063595.

XX 28-APR-1997; 97US-0044428.

XX 20-MAY-1997; 97US-0047222.

XX 11-JUN-1997; 97US-0049477.

XX 11-JUN-1997; 97US-0049756.

XX 18-JUN-1997; 97US-0049914.

XX (SUGE-) SUGEN INC.

XX App H, Clary D, Courtneidge SA, Hui TH, Jallal B;

PI Markby D, Onrust S, Peles E, Plowman GD;

DR WPI; 1999-009434/01.

DR P-PSDB; AAW89249.

XX New nucleic acid encoding specific protein tyrosine phosphatases -  
 PT useful for identifying specific modulators for treatment and  
 PT prevention of cancer and neurodegenerative disease

XX Claim 2; Page 147; 193pp; English.

XX The present invention describes isolated, enriched or purified nucleic  
 CC acids encoding PTP04, SAD, PTP05, PTP10, ALP and ALK-7 proteins. The  
 CC present sequence encodes mouse PTP05. The above proteins, other than  
 CC ALK-7, are protein tyrosine phosphatases (PTPs) and are used to identify  
 CC substances that modulate their activity (i.e. agonists and antagonists,  
 CC including NBP) in vivo or in vitro. These substances are used to treat  
 CC or prevent diseases associated with abnormal signal transduction  
 CC pathways that involve the proteins, particularly cancer (e.g. leukaemia  
 CC and lymphoma), while modulators of ALK-7 (which is a type I receptor  
 CC serine/threonine kinase) are used to promote neuronal survival,  
 CC particularly for treating Alzheimer's, Parkinson's or Huntington's  
 CC diseases. Nucleic acid fragments of the polynucleotides encoding the  
 CC proteins can be used as probes to identify and clone related sequences;  
 CC to detect protein-encoded RNA; to generate transgenic animals and in

CC gene therapy (optionally after mutation). Ab are used to determine the  
 CC proteins.

SQ Sequence 1785 BP; 527 A; 361 C; 408 G; 489 T; 0 other;

Query Match 87.5%; Score 280; DB 20; Length 1785;  
 Best Local Similarity 92.2%; Pred. No. 2.4e-76;  
 Matches 295; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

Qy 1 GAAATATATTGTAATGTTATTGCTATGATAACCCAGAGATAGAGGTGGAGTTATCAAG 60  
 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 949 gaaataattgcaatgttcctcatgataaccagagagatagaatgtggagttatcaag 1008

Qy 51 TGTTCAGATTACTGGCCGCTTCTCTGAAGAGAGCCTTTGGAAATTCAAACACTTTTCATGTC 120  
 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 1009 tgttacagttactggccattctctgaagagagcctttggaattcgaacacttttagtgc 1068

Qy 121 CTCTGGAGAACTTTCAGATAACTCAGTATTTTGTTCATCCGAATATTTCAAAATGTTGAAG 180  
 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 1069 ttctggagaccttctcatgtaactcaatatttcacccgttcagtagtattcagattgtgaag 1128

Qy 181 AAGTCCACAGAAAGAGTCACTCTGTAAACACATTGCAAGTTTCATCAAAATGCCAGACCAT 240  
 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 1129 aagtcacagaaagagagccaatgtgtaaacacattgcagttccaccagagccagaccat 1188

Qy 241 GGCACTCTGCTCAGTAGATTTTTCATCAATATGTCCTTATGTGAGGAAGAGCCAC 300  
 |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 Db 1189 ggcactctgctcagcagattttttcataaaatgctccgttatgtgaggaagagccac 1248

Qy 301 ATTACAGAGCCCTCCTTGT 320  
 |||||||||||||||||||||||

Db 1249 attacaggagccctccttgt 1268

## RESULT 4

AAV81745

ID AAV81745 standard; cDNA; 1896 BP.

AC AAV81745;

DT 10-MAR-1999 (first entry)

DE Mouse PTP05 isoform #1 encoding cDNA.

XX PTP04; PTP05; PTP10; SAD; ALP; ALK-7; protein tyrosine phosphatase;  
 KW type I receptor serine/threonine kinase; cancer; leukaemia; lymphoma;  
 KW neurodegenerative disease; neuronal survival; Alzheimer's disease;  
 KW Parkinson's disease; Huntington's disease; ss.

OS Mus sp.

XX Key Location/Qualifiers  
 FH 199..1590  
 FT CDS /\*tag= a  
 XX

PN W09849317-A2.

XX 05-NOV-1998.

XX 27-APR-1998; 98WO-US08439.

XX 23-OCT-1997; 97US-0063595.

XX 28-APR-1997; 97US-0044428.

XX 20-MAY-1997; 97US-0047222.

XX 11-JUN-1997; 97US-0049477.

XX 11-JUN-1997; 97US-0049756.

XX 18-JUN-1997; 97US-0049914.

XX (SUGE-) SUGEN INC.

XX App H, Clary D, Courtneidge SA, Hui TH, Jallal B;

PI Markby D, Onrust S, Peles E, Plowman GD;





```
Db 3848 gcaatgacatcagaaaattgtttcttcggaatggaccatcagagatttcacagtgaaaaat 3907
QY 181 AAGTCCACAGAAAGAGTCACTCTGTAAAACACTTGCAGTTCAATAAGCCAGACCAT 240
Db 3908 atccagacaagtgcagtcaccctctgcagacagttccattccctcctgcgcagaccac 3967
QY 241 GGCACCTCCTGCCTCAGTAGATTTTTCATCAAAAT 274
Db 3968 ggtgtccgcagaccactgcactgcctcactcaact 4001

RESULT 8
AAF58252/C
ID AAF58252 standard; DNA; 936 BP.
XX
AC AAF58252;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D1835.
XX
KW Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
XX
OS Synthetic.
XX
PN WO200107665-A2.
XX
PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000WO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
XX
PR 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI Umek RM;
XX
XX WPI; 2001-159728/16.
XX
DR Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface
XX
PS Example 6; Page 127; 159pp; English.
XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SQ Sequence 936 BP; 4 A; 139 C; 10 G; 7 T; 776 other;

Query Match 14.1%; Score 45; DB 22; Length 936;
Best Local Similarity 1.0%; Pred. No. 0.00036;
Matches 3; Conservative 185; Mismatches 115; Indels 0; Gaps 0;

QY 2 AAAATAATTCGTAATGTATTGCTATGATACACAGAGATAGAGTGGAGTTATCAAGT 61
Db 778 WWWWWW...
QY 62 GTTGCAGTTACTGGCCGGTTCTCTGAAGAGCGCTTTGGAATTCAAACACTTTTCATGTC 121
Db 718 WWWWWW...
QY 122 TTCTGGAGAACTTTCAGATAACTCAGTATTTTGTCTATCCGAATATTTCAAAATTTGTAAGA 181
Db 658 WWWWWW...
599
```

```
QY 182 AGTCCACAGAAAGAGTCACTCTGTAAAACACTTGCAGTTCAATAAGCCAGACCATG 241
Db 598 WWWWWW...
QY 242 GCACTCCTGCCTCAGTAGATTTTTCATCAAAATATGTCCCTTATGTCAGGAAGAGCCACA 301
Db 538 WWWWWW...
QY 302 TTA 304
Db 478 WWW 476

RESULT 9
AAF58254/C
ID AAF58254 standard; DNA; 936 BP.
XX
AC AAF58254;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D1875.
XX
KW Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
XX
OS Synthetic.
XX
PN WO200107665-A2.
XX
PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000WO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
XX
PR 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI Umek RM;
XX
XX WPI; 2001-159728/16.
XX
DR Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface
XX
PS Example 6; Page 127; 159pp; English.
XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SQ Sequence 936 BP; 4 A; 144 C; 7 G; 5 T; 776 other;

Query Match 14.1%; Score 45; DB 22; Length 936;
Best Local Similarity 1.0%; Pred. No. 0.00036;
Matches 3; Conservative 185; Mismatches 115; Indels 0; Gaps 0;

QY 2 AAAATAATTCGTAATGTATTGCTATGATACACAGAGATAGAGTGGAGTTATCAAGT 61
Db 778 WWWWWW...
QY 52 GTTGCAGTTACTGGCCGGTTCTCTGAAGAGCGCTTTGGAATTCAAACACTTTTCATGTC 121
Db 718 WWWWWW...
QY 122 TTCTGGAGAACTTTCAGATAACTCAGTATTTTGTCTATCCGAATATTTCAAAATTTGTAAGA 181
Db 659
```

```
Db 658 WWWWWW 599
QY 182 AGTCACAGAAAGAGTCACCTCTGTAAACACCTTCAGTTCATCAAAATGGCCAGACCATG 241
Db 598 WWWWWW 539
QY 242 GCACCTCCCTGCCTCAGTAGATTTTCATCAAAATATGTCGTTATCTGAGGAGAGCCACA 301
Db 538 WWWWWW 479
QY 302 TTA 304
Db 478 WWW 476

RESULT 10
AAF58257/C
ID AAF58257 standard; DNA; 936 BP.
XX
AC AAF58257;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D1934.
XX
KW Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
XX
OS Synthetic.
XX
PN WO200107665-A2.
XX
PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000WO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
PR 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI Umek RM;
XX
DR WPI; 2001-159728/16.
XX
PT Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface
XX
PS Example 6; Page 127; 159pp; English.
XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SQ Sequence 936 BP; 5 A; 142 C; 7 G; 6 T; 776 other;

Query Match 14.1%; Score 45; DB 22; Length 936;
Best Local Similarity 1.0%; Pred. No. 0.00036;
Matches 3; Conservative 185; Mismatches 115; Indels 0; Gaps 0;

QY 2 AAAATAATTGTAATCTTATGCTATGATAACACAGAGATAGAAGTGGAGTTATCAAGT 61
Db 778 WWWWWW 719
QY 62 GTTCAGTTACTGGCCGCTTCTCTGAAGAGCCCTTTGGAATTCACACACTTTTCATGTC 121
Db 718 WWWWWW 659
```

```
QY 122 TTCTGGAGAACTTTCAGATAACTCAGTATTTTGTCTATCGAATATTTCAAATTTGGAAGA 181
Db 658 WWWWWW 599
QY 182 AGTCACAGAAAGAGTCACCTCTGTAAACACCTTCAGTTCATCAAAATGGCCAGACCATG 241
Db 598 WWWWWW 539
QY 242 GCACCTCCCTGCCTCAGTAGATTTTCATCAAAATATGTCGTTATCTGAGGAGAGCCACA 301
Db 538 WWWWWW 479
QY 302 TTA 304
Db 478 WWW 476

RESULT 11
ZAF58259/C
ID AAF58259 standard; DNA; 936 BP.
XX
AC AAF58259;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D2004.
XX
KW Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
XX
OS Synthetic.
XX
PN WO200107665-A2.
XX
PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000WO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
PR 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI Umek RM;
XX
DR WPI; 2001-159728/16.
XX
PT Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface
XX
PS Example 6; Page 128; 159pp; English.
XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SQ Sequence 936 BP; 6 A; 138 C; 8 G; 8 T; 776 other;

Query Match 14.1%; Score 45; DB 22; Length 936;
Best Local Similarity 1.0%; Pred. No. 0.00036;
Matches 3; Conservative 185; Mismatches 115; Indels 0; Gaps 0;

QY 2 AAAATAATTGTAATCTTATGCTATGATAACACAGAGATAGAAGTGGAGTTATCAAGT 61
Db 778 WWWWWW 719
QY 62 GTTCAGTTACTGGCCGCTTCTCTGAAGAGCCCTTTGGAATTCACACACTTTTCATGTC 121
```

[illegible][illegible]



PT Nucleic acids containing electron-transfer group, useful as labels in  
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on  
PT a single surface

XX  
PS  
XX  
XX  
XX

Example 6; Page 127; 159pp; English.

CC The present invention relates to a composition comprising two nucleic  
CC acids each containing an electron-transfer group (ETM) having  
CC different redox potentials. The invention is used for electronic  
CC detection of nucleic acids, especially of substitutions (mismatches)  
CC and single-nucleotide polymorphisms, e.g. for genotyping,  
CC monitoring gene expression.

XX  
SQ Sequence 936 BP; 4 A; 139 C; 10 G; 7 T; 776 other;

Query Match	13.9%;	Score 44.4;	DB 22;	Length 936;
Best Local Similarity	3.8%;	Pred. No. 0.00055;		
Matches 12;	Conservative 178;	Mismatches 124;	Indels 0;	Gaps 0;

  

Qy	3	AAATAATGTAATGTTATGCTATGATAACCCAGAGATAGAAGGTGGAGTTATCAAGTG	62
Db	495	www	554

  

Qy	63	TTGCAGTTACTGGCCCGTTTCTCTGAAGGAGCCCTTTGGAATTCAAACACTTTCATGTCCT	122
Db	555	www	614

  

Qy	123	TCTGGAGAACTTTTCAGATAACTCAGTATTTTGTCTATCCGAATATTCAAATTTGTGAAGAA	182
Db	615	www	674

  

Qy	183	GTCCACAGGAAAGAGTCACCTCTGTAAACACTTGCAGTTTCATCAAAATGCCAGACCATGG	242
Db	675	www	734

  

Qy	243	CACCTCTGCCTCAGTAGATTTTTCATCAATATGTCCTTATGTGAGGAGAGCCACAT	302
Db	735	www	794

  

Qy	303	TACAGGACCCCTCC	316
Db	795	cccccccccccccc	808

Search completed: March 30, 2002, 05:19:42  
Job time: 3410 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2002, 06:07:33 ; Search time 175.91 Seconds  
(without alignments)  
1559.571 Million cell updates/sec

Title: us-09-095-478a-4  
Perfect score: 320  
Sequence: 1 GAAATAATGTGTAATGTTAT.....ATTACAGGACCCCTCTTGT 320

Scoring table: OLIGO\_NUC  
Gapop 60.0 , Gapext 60.0

Searched: 930621 seqs, 428662619 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : N\_Geneseq\_1101.\*

1: /SID22/gcgdata/geneseq/geneseq/NA1980.DAT.\*  
2: /SID22/gcgdata/geneseq/geneseq/NA1981.DAT.\*  
3: /SID22/gcgdata/geneseq/geneseq/NA1982.DAT.\*  
4: /SID22/gcgdata/geneseq/geneseq/NA1983.DAT.\*  
5: /SID22/gcgdata/geneseq/geneseq/NA1984.DAT.\*  
6: /SID22/gcgdata/geneseq/geneseq/NA1985.DAT.\*  
7: /SID22/gcgdata/geneseq/geneseq/NA1986.DAT.\*  
8: /SID22/gcgdata/geneseq/geneseq/NA1987.DAT.\*  
9: /SID22/gcgdata/geneseq/geneseq/NA1988.DAT.\*  
10: /SID22/gcgdata/geneseq/geneseq/NA1989.DAT.\*  
11: /SID22/gcgdata/geneseq/geneseq/NA1990.DAT.\*  
12: /SID22/gcgdata/geneseq/geneseq/NA1991.DAT.\*  
13: /SID22/gcgdata/geneseq/geneseq/NA1992.DAT.\*  
14: /SID22/gcgdata/geneseq/geneseq/NA1993.DAT.\*  
15: /SID22/gcgdata/geneseq/geneseq/NA1994.DAT.\*  
16: /SID22/gcgdata/geneseq/geneseq/NA1995.DAT.\*  
17: /SID22/gcgdata/geneseq/geneseq/NA1996.DAT.\*  
18: /SID22/gcgdata/geneseq/geneseq/NA1997.DAT.\*  
19: /SID22/gcgdata/geneseq/geneseq/NA1998.DAT.\*  
20: /SID22/gcgdata/geneseq/geneseq/NA1999.DAT.\*  
21: /SID22/gcgdata/geneseq/geneseq/NA2000.DAT.\*  
22: /SID22/gcgdata/geneseq/geneseq/NA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	320	100.0	320	AAV81747	Rat PTP10 encoding
2	50	15.6	1692	AAV81746	Mouse PTP05 isofor
3	50	15.6	1785	AAV81744	Mouse PTP05 encodi
4	50	15.6	1896	AAV81745	Mouse PTP05 isofor
5	33	10.3	33	AAV81754	Human PTP04 RACE P
6	33	10.3	33	AAV81755	Human PTP04 RACE P
7	21	6.6	477	AAH94038	Human foetal cDNA,
8	21	6.6	1184	AAH94038	Human foetal cDNA,
9	18	5.6	192	AAH21420	Human gene signatu
10	18	5.6	304	AAA44583	Human secreted exp
11	18	5.6	332	AAH50777	Human tumour assoc

12	18	5.6	1398	20	AAV81747	Human HDRC DNA. H
13	17	5.3	182	22	AAV81747	Probe #22802 used
14	17	5.3	528	22	AAV81747	Human lung tumour
15	17	5.3	538	22	AAV81747	Probe #9737 used t
16	17	5.3	542	22	AAV81747	Human CDNA clone (
17	17	5.3	587	22	AAV81747	Human CDNA clone (
18	17	5.3	696	21	AAV81747	Aspergillus oryzae
19	17	5.3	718	22	AAV81747	Human EST-derived
20	17	5.3	749	22	AAV81747	Human CDNA clone (
21	17	5.3	785	11	AAV81747	Human arginase cDN
22	17	5.3	878	21	AAV81747	Lung cancer associ
23	17	5.3	931	21	AAV81747	Human secreted pro
24	17	5.3	931	22	AAV81747	Human secreted pro
25	17	5.3	966	11	AAV81747	Human arginase cod
26	17	5.3	969	19	AAV81747	Human arginase enc
27	17	5.3	972	19	AAV81747	Rat arginase encod
28	17	5.3	1365	11	AAV81747	Human arginase cDN
29	17	5.3	1437	22	AAV81747	Human liver cell s
30	17	5.3	1563	21	AAV81747	Arabidopsis thalia
31	17	5.3	1736	21	AAV81747	Human secreted pro
32	17	5.3	1772	22	AAV81747	Human CDNA sequenc
33	17	5.3	1848	21	AAV81747	Arabidopsis thalia
34	17	5.3	2922	22	AAV81747	Human CDNA sequenc
35	17	5.3	3299	22	AAV81747	Human CDNA sequenc
36	17	5.3	5558	19	AAV81747	Streptococcus pneu
37	17	5.3	138169	21	AAV81747	Human adenosine re
38	17	5.3	141589	21	AAV81747	Human ELAM-1 polyn
39	17	5.3	141589	21	AAV81747	Human low adenosin
40	17	5.3	141589	21	AAV81747	Human low adenosin
41	17	5.3	141589	21	AAV81747	Human adenosine re
42	17	5.3	141589	21	AAV81747	Human adenosine re
43	17	5.3	143068	21	AAV81747	Human low adenosin
44	17	5.3	143068	21	AAV81747	Human low adenosin
45	17	5.3	143068	21	AAV81747	Human adenosine re

#### ALIGNMENTS

RESULT : 1

AAV81747

ID AAV81747 standard; cDNA; 320 BP.

XX AAV81747;

AC AAV81747;

XX ..

DT 10-MAR-1999 (first entry)

XX ..

DE Rat PTP10 encoding cDNA.

XX ..

KW PTP04; PTP05; PTP10; SAD; ALP; ALK-7; protein tyrosine phosphatase;

KW type I receptor serine/threonine kinase; cancer; leukaemia; lymphoma;

KW neurodegenerative disease; neuronal survival; Alzheimer's disease;

KW Parkinson's disease; Huntington's disease; ss.

XX ..

OS Rattus sp.

XX ..

PN WO9849317-A2.

XX ..

PD 05-NOV-1998.

XX ..

PF 27-APR-1998; 98WO-US08439.

XX ..

PR 23-OCT-1997; 97US-0063595.

PR 28-APR-1997; 97US-0044428.

PR 20-MAY-1997; 97US-0047222.

PR 11-JUN-1997; 97US-0049477.

PR 11-JUN-1997; 97US-0049756.

PR 18-JUN-1997; 97US-0049914.

XX ..

PA (SUGEN-) SUGEN INC.

XX ..

PI App H, Clary D, Courtneidge SA, Hui TH, Jallal B;

PI Markby D, Onrust S, Peles E, Prowman GD;



[illegible]

```

XX 10-MAR-1999 (first entry)
XX Mouse PTP05 isoform #1 encoding CDNA.
XX
XX PTP04; PTP05; PTP10; SAD; ALP; ALK-7; protein tyrosine phosphatase;
XX type I receptor serine/threonine kinase; cancer; leukaemia; lymphoma;
XX neurodegenerative disease; neuronal survival; Alzheimer's disease;
XX Parkinson's disease; Huntington's disease; ss.
XX
XX Muls sp.
XX
XX Key Location/Qualifiers
XX FT 199..1590
XX FT ; /*tag= a
XX
XX WO9849317-A2.
XX
XX 05-NOV-1998..
XX
XX 27-APR-1998; 98WO-US08439..
XX
XX 23-OCT-1997; 97US-0063595.
XX 28-APR-1997; 97US-0044428.
XX 20-MAY-1997; 97US-0047222.
XX 11-JUN-1997; 97US-0049477.
XX 11-JUN-1997; 97US-0049756.
XX 18-JUN-1997; 97US-0049914.
XX
XX (SUGB-) SUGEN INC.
XX
XX App H, Clary D, Courtneidge SA, Hui TH, Jallal B;
XX Markby D, Onrust S, Peles E, Plozman GD;
XX
XX WPI: 1999-002434/01.
XX P-PSDB; AAW89250.
XX
XX New nucleic acid encoding specific protein tyrosine phosphatases -
XX useful for identifying specific modulators for treatment and
XX prevention of cancer and neurodegenerative disease
XX
XX Claim 2; Page 147-148; 193pp; English.
XX
XX The present invention describes isolated, enriched or purified nucleic
XX acids encoding PTP04, SAD, PTP05, PTP10, ALP and ALK-7 proteins. The
XX present sequence encodes mouse PTP05. The above proteins, other than
XX ALK-7, are protein tyrosine phosphatases (PTPs) and are used to identify
XX substances that modulate their activity (i.e. agonists and antagonists,
XX including NBP) in vivo or in vitro. These substances are used to treat
XX or prevent diseases associated with abnormal signal transduction
XX pathways that involve the proteins, particularly cancer (e.g. leukaemia
XX and lymphoma), while modulators of ALK-7 (which is a type I receptor
XX serine/threonine kinase) are used to promote neuronal survival,
XX particularly for treating Alzheimer's, Parkinson's or Huntington's
XX diseases. Nucleic acid fragments of the polynucleotides encoding the
XX proteins can be used as probes to identify and clone related sequences;
XX to detect protein-encoded RNA; to generate transgenic animals and in
XX gene therapy (optionally after mutation). Ab are used to determine the
XX proteins.
XX
XX Sequence 1896 BP; 572 A; 372 C; 427 G; 525 T; 0 other;
XX
XX
XX Query Match 15.6%; Score 50; DB 20; Length 1896;
XX Best Local Similarity 100.0%; Pred. No. 1.9e-16;
XX Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps
XX
XX 271 AAATATGTCGTTATGTGAGGAAGACCCACATTACAGGACCCCTCCTTCT 320
XX |||||
XX 1330 aaatagtctcgttatgtgagggaagccacattacaggagccctcttctgt 1379
XX
XX RESULT 5

```

AAV81754  
ID AAV81754 standard; DNA; 33 BP.

XX AC AAV81754;

XX DT 10-MAR-1999 (first entry)

XX DE Human PTP04 RACE primer 6599.

XX

KW PTP04; PTP05; PTP10; SAD; ALP; ALK-7; protein tyrosine phosphatase;  
KW type I receptor serine/threonine kinase; cancer; leukaemia; lymphoma;  
KW neurodegenerative disease; neuronal survival; Alzheimer's disease;  
KW Parkinson's disease; Huntington's disease; PCR primer; ss.

XX

OS Synthetic.

OS Homo sapiens.

XX

PN WO9849317-A2.

XX

PD 05-NOV-1998.

XX

XX PF 27-APR-1998; 98WO-US08439.

XX

PR 23-OCT-1997; 97US-0063595.

PR

PR 28-APR-1997; 97US-0044428.

PR

PR 20-MAY-1997; 97US-0047222.

PR

PR 11-JUN-1997; 97US-0049477.

PR

PR 11-JUN-1997; 97US-0049756.

PR

PR 18-JUN-1997; 97US-0049914.

XX

XX (SUGEN-) SUGEN INC.

XX

PI App H, Clary D, Courtneidge SA, Hui TH, Jallal B;

PI Markby D, Onrust S, Peles E, Plowman GD;

XX

XX WPI; 1999-009434/01.

XX

XX New nucleic acid encoding specific protein tyrosine phosphatases -

PT useful for identifying specific modulators for treatment and

PT prevention of cancer and neurodegenerative disease

XX

XX Example 1; Page 74; 193pp; English.

XX

XX The present invention describes isolated, enriched or purified nucleic

CC acids encoding PTP04, SAD, PTP05, PTP10, ALP and ALK-7 proteins. The

CC above proteins, other than ALK-7, are protein tyrosine phosphatases

CC (PTPs) and are used to identify substances that modulate their activity

CC (i.e. agonists and antagonists, including NBP) in vivo or in vitro.

CC These substances are used to treat or prevent diseases associated with

CC abnormal signal transduction pathways that involve the proteins,

CC particularly cancer (e.g. leukaemia and lymphoma), while modulators of

CC ALK-7 (which is a type I receptor serine/threonine kinase) are used to

CC promote neuronal survival, particularly for treating Alzheimer's,

CC Parkinson's or Huntington's diseases. Nucleic acid fragments of the

CC polynucleotides encoding the proteins can be used as probes to identify

CC and clone related sequences; to detect protein-encoded RNA; to generate

CC transgenic animals and in gene therapy (optionally after mutation). Ab

CC are used to determine the proteins. The present sequence represents a

CC RACE primer for human PTP04.

XX

SQ Sequence 33 BP; 10 A; 8 C; 9 G; 6 T; 0 other;

XX

Query Match 10.3%; Score 33; DB 20; Length 33;

Best Local Similarity 100.0%; Pred. No. 1.3e-07;

Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 279 CCGTTATGTCGAGGAGAGCCACATTACAGACC 311

Db 1 cccgttatgtgaggaagagccacattacagacc 33

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

AAV81755/C

XX ID AAV81755 standard; DNA; 33 BP.

XX AC AAV81755;

XX DT 10-MAR-1999 (first entry)

XX DE Human PTP04 RACE primer 6600.

XX

KW PTP04; PTP05; PTP10; SAD; ALP; ALK-7; protein tyrosine phosphatase;

KW type I receptor serine/threonine kinase; cancer; leukaemia; lymphoma;

KW neurodegenerative disease; neuronal survival; Alzheimer's disease;

KW Parkinson's disease; Huntington's disease; PCR primer; ss.

XX

OS Synthetic.

OS Homo sapiens.

XX

PN WO9849317-A2.

XX

PD 05-NOV-1998.

XX

XX PF 27-APR-1998; 98WO-US08439.

XX

PR 23-OCT-1997; 97US-0063595.

PR

PR 28-APR-1997; 97US-0044428.

PR

PR 20-MAY-1997; 97US-0047222.

PR

PR 11-JUN-1997; 97US-0049477.

PR

PR 11-JUN-1997; 97US-0049756.

PR

PR 18-JUN-1997; 97US-0049914.

XX

XX (SUGEN-) SUGEN INC.

XX

PI App H, Clary D, Courtneidge SA, Hui TH, Jallal B;

PI Markby D, Onrust S, Peles E, Plowman GD;

XX

XX WPI; 1999-009434/01.

XX

XX New nucleic acid encoding specific protein tyrosine phosphatases -

PT useful for identifying specific modulators for treatment and

PT prevention of cancer and neurodegenerative disease

XX

XX Example 1; Page 74; 193pp; English.

XX

XX The present invention describes isolated, enriched or purified nucleic

CC acids encoding PTP04, SAD, PTP05, PTP10, ALP and ALK-7 proteins. The

CC above proteins, other than ALK-7, are protein tyrosine phosphatases

CC (PTPs) and are used to identify substances that modulate their activity

CC (i.e. agonists and antagonists, including NBP) in vivo or in vitro.

CC These substances are used to treat or prevent diseases associated with

CC abnormal signal transduction pathways that involve the proteins,

CC particularly cancer (e.g. leukaemia and lymphoma), while modulators of

CC ALK-7 (which is a type I receptor serine/threonine kinase) are used to

CC promote neuronal survival, particularly for treating Alzheimer's,

CC Parkinson's or Huntington's diseases. Nucleic acid fragments of the

CC polynucleotides encoding the proteins can be used as probes to identify

CC and clone related sequences; to detect protein-encoded RNA; to generate

CC transgenic animals and in gene therapy (optionally after mutation). Ab

CC are used to determine the proteins. The present sequence represents a

CC RACE primer for human PTP04.

XX

SQ Sequence 33 BP; 6 A; 9 C; 8 G; 10 T; 0 other;

XX

Query Match 10.3%; Score 33; DB 20; Length 33;

Best Local Similarity 100.0%; Pred. No. 1.3e-07;

Matches 33; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 279 CCGTTATGTCGAGGAGAGCCACATTACAGACC 311

Db 33 CCGTTATGTCGAGGAGAGCCACATTACAGACC 1

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

AAH94038  
ID AAH94038 standard; cDNA; 477 BP.  
XX  
AC AAH94038;  
XX  
DT 05-OCT-2001 (first entry)  
XX  
DE Human foetal cDNA, SEQ ID NO: 567.  
XX  
KW Human; foetal protein; cytostatic; immunosuppressive; immunostimulant;  
KW neotropic; neuroprotective; thrombolytic; osteopathic; antiinflammatory;  
KW gene therapy; antisense therapy; cancer; immune disorder;  
KW growth disorder; osteoporosis; thrombolytic disorder;  
KW nervous system disorder; inflammation; expressed sequence tag; EST; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200155339-A2.  
XX  
PD 02-AUG-2001.  
XX  
PF 25-JAN-2001; 2001WO-US02723.  
XX  
PR 25-JAN-2000; 2000US-0491404.  
PR 15-SEP-2000; 2000US-0663870.  
PR 06-NOV-2000; 2000US-0707351.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Yeung G, Ford JE, Boyle BJ, Arterburn MC, Drmanac RA, Tang YT;  
PI Liu C, Asundi V, Zhou P, Werhman T;  
XX  
DR WPI; 2001-465571/50.  
DR P-PSDB; AAM06363.  
XX  
PT Novel fetal proteins useful for the treatment and diagnosis of diseases  
PT associated with dysfunction of the protein e.g. cancers, immune  
PT disorders, growth disorders, thrombolytic disorders, nervous system  
PT disorders and inflammation -  
XX  
PS Claim 1; Page 385; 715pp; English.  
XX  
CC The invention relates to novel foetal polypeptides encoded by  
CC polynucleotides comprising one of 477 sequences fully defined in the  
CC specification. The foetal polynucleotides and polypeptides are  
CC useful in the treatment and diagnosis of diseases such as cancers,  
CC immune disorders, growth disorders (e.g. osteoporosis), thrombolytic  
CC disorders, nervous system disorders and inflammation. The present  
CC sequence was assembled using an expressed sequence tag (EST) found  
CC to be expressed in human foetal tissue cDNA libraries as the seed.  
XX  
SQ Sequence 477 BP; 128 A; 90 C; 78 G; 180 T; 1 other;  
  
Query Match 6.6%; Score 21; DB 22; Length 477;  
Best Local Similarity 100.0%; Pred. No. 0.27;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 103 TTCAAACACTTTCATGTCCTT 123  
Db 281 ttcaaacactttcatgctctt 301  
  
RESULT 8  
AAH94456  
ID AAH94456 standard; cDNA; 1184 BP.  
XX  
AC AAH94456;  
XX  
DT 05-OCT-2001 (first entry)  
XX  
DE Human foetal cDNA, SEQ ID NO: 1143.  
XX

KW Human; foetal protein; cytostatic; immunosuppressive; immunostimulant;  
KW neotropic; neuroprotective; thrombolytic; osteopathic; antiinflammatory;  
KW gene therapy; antisense therapy; cancer; immune disorder;  
KW growth disorder; osteoporosis; thrombolytic disorder;  
KW nervous system disorder; inflammation; expressed sequence tag; EST; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200155339-A2.  
XX  
PD 02-AUG-2001.  
XX  
PF 25-JAN-2001; 2001WO-US02723.  
XX  
PR 25-JAN-2000; 2000US-0491404.  
PR 15-SEP-2000; 2000US-0663870.  
PR 06-NOV-2000; 2000US-0707351.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Yeung G, Ford JE, Boyle BJ, Arterburn MC, Drmanac RA, Tang YT;  
PI Liu C, Asundi V, Zhou P, Werhman T;  
XX  
DR WPI; 2001-465571/50.  
DR P-PSDB; AAM06781.  
XX  
PT Novel fetal proteins useful for the treatment and diagnosis of diseases  
PT associated with dysfunction of the protein e.g. cancers, immune  
PT disorders, growth disorders, thrombolytic disorders, nervous system  
PT disorders and inflammation -  
XX  
PS Example 3; Page 641; 715pp; English.  
XX  
CC The invention relates to novel foetal polypeptides encoded by  
CC polynucleotides comprising one of 477 sequences fully defined in the  
CC specification. The foetal polynucleotides and polypeptides are  
CC useful in the treatment and diagnosis of diseases such as cancers,  
CC immune disorders, growth disorders (e.g. osteoporosis), thrombolytic  
CC disorders, nervous system disorders and inflammation. The present  
CC sequence is a full length cDNA which was assembled using expressed  
CC sequence tags (ESTs) found to be expressed in human foetal tissue  
CC cDNA libraries as seeds.  
XX  
SQ Sequence 1184 BP; 320 A; 269 C; 244 G; 351 T; 0 other;  
  
Query Match 6.6%; Score 21; DB 22; Length 1184;  
Best Local Similarity 100.0%; Pred. No. 0.28;  
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 103 TTCAAACACTTTCATGTCCTT 123  
Db 273 ttcaaacactttcatgctctt 293  
  
RESULT 9  
AAAT21420  
ID AAAT21420 standard; cDNA to mRNA; 192 BP.  
XX  
AC AAAT21420;  
XX  
DT 02-AUG-1996 (first entry)  
XX  
DE Human gene signature HUMGS02790.  
XX  
KW Gene signature; messenger RNA; mRNA; relative abundance; frequency;  
KW human; cloning; mapping; non-biased library; diagnosis; detection;  
KW cell typing; abnormal cell function; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO9514772-A1.  
XX

PD 01-JUN-1995.  
 XX  
 PF 11-NOV-1994; 94WO-JP01916.  
 XX  
 PR 12-NOV-1993; 93JP-0355504.  
 XX  
 PA (MATS/) MATSUBARA K.  
 PA (OKUB/) OKUBO K.  
 XX  
 PI Matsubara K, Okubo K;  
 PI WPI; 1995-206931/27.  
 DR  
 XX  
 XX Identifying gene signatures in 3'-directed human cDNA library - e.g.  
 PT for diagnosis of abnormal cell function, by preparing cDNA that  
 PT reflects relative abundance of corresp. mRNA in specific human  
 PT tissues  
 XX  
 XX  
 PS Claim 1; Page 882; 2245pp; Japanese.  
 XX  
 CC A single-stranded DNA (or its complementary strand or the corresp.  
 CC double-stranded DNA) which comprises one of the 7837 "GS" sequences  
 CC given in AAT19001-T26837 and which is able to hybridise to part of  
 CC human genomic DNA, cDNA or mRNA is claimed. The GS (Gene Signature)  
 CC sequences were obtained from 3'-directed cDNA libraries prepared  
 CC from various human tissues; synthesis of cDNA was initiated from the  
 CC 3'-end of mRNA by using poly(T) as the sole primer. Since the 3'-  
 CC untranslated sequence is unique to a particular mRNA species, almost  
 CC all the 3'-oriented cDNAs hybridise with specific mRNAs. Each library  
 CC is constructed so as to reflect accurately the relative abundance of  
 CC different mRNAs in the particular tissue from which it was derived.  
 CC The appearance frequency of a given GS in a cDNA library can be  
 CC determined (esp. using primers and probes derived from the GS  
 CC sequences) as a means of diagnosing abnormal cell function or for  
 CC recognising different cell types.  
 XX  
 SQ Sequence 192 BP; 67 A; 25 C; 21 G; 77 T; 2 other;  
 XX  
 Query Match 5.6%; Score 18; DB 16; Length 192;  
 Best Local Similarity 100.0%; Pred. No. 9.9;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 XX  
 QY 1 GAAATTAATTTGTAATGTT 18  
 DB |||||  
 DB 111 gaaataaattgtaattgt 128  
 XX  
 RESULT 10  
 AAA44583  
 ID AAA44583 standard; cDNA; 304 BP.  
 XX  
 AC AAA44583;  
 XX  
 XX 21-AUG-2000 (first entry)  
 DT Human secreted expressed sequence tag SEQ ID NO:1158.  
 DE  
 XX Human; mouse; chicken; rat; secreted expressed sequence tag; SEST;  
 KW expressed sequence tag; EST; probe; chemotactic; proliferative;  
 KW immunomodulatory; haematopoietic; chemokinetic; analgesic; haemostatic;  
 KW thrombolytic; antiinflammatory; cytostatic; antibacterial; antifungal;  
 KW antiviral; antidiabetic; antiasthmatic; vulnary; antiparkinsonian;  
 KW antitumor; osteoprotective; neuroprotective; neurotropic; antipsoriatic;  
 KW cerebroprotective; anticonvulsant; antidepressant; gene therapy;  
 KW vaccine; autoimmune disorder; multiple sclerosis; allergic condition;  
 KW insulin dependent diabetes; asthma; myeloid cell deficiency; ulcer;  
 KW lymphoid cell deficiency; burn; osteoporosis; osteoarthritis;  
 KW central nervous system disorder; Alzheimer's disease; stroke;  
 KW Parkinson's disease; Huntington's disease; coagulation disorder;  
 KW haemophilia; thrombosis; inflammatory disorder; Crohn's disease;  
 KW tumour; infection; depression; psoriasis; ss.  
 XX

OS Homo sapiens.  
 XX WC200021991-A1.  
 PN  
 XX 20-APR-2000.  
 PD  
 XX 15-OCT-1999; 99WO-US24206.  
 PF  
 XX 15-OCT-1998; 98US-0104436.  
 PR  
 XX (GEMV ) GENETICS INST INC.  
 PA  
 XX Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;  
 PI Merberg D, Treacy M, Bowman MK;  
 PI WPI; 2000-317938/27.  
 XX  
 XX Isolated polynucleotides, and encoded proteins, comprising secreted  
 PT expressed sequence tags (SESTs), useful for treating various disorders  
 PT such as autoimmune, infectious, and central nervous system disorders -  
 PT  
 XX Claim 1; Page 500; 803pp; English.  
 XX  
 CC AAA43426 to AAA45925 represent specifically claimed secreted expressed  
 CC sequence tags (SESTs), isolated from human, mouse, chicken and rat  
 CC tissue sources. The SESTs can have a range of activities depending on  
 CC the tissues they were isolated from. The activities include:  
 CC chemokinetic; proliferative; immunomodulatory; haematopoietic;  
 CC chemotactic; analgesic; haemostatic; thrombolytic; antiinflammatory;  
 CC cytostatic; antibacterial; antifungal; antiviral; antidiabetic;  
 CC antiasthmatic; vulnary; antitumor; osteopathic; neuroprotective;  
 CC neurotropic; antiparkinsonian; antipsoriatic; cerebroprotective;  
 CC anticonvulsant; and antidepressant. The SESTs can be used for gene  
 CC therapy and in vaccines. The SESTs are useful as probes for the  
 CC identification and isolation of full-length cDNAs and genomic DNA  
 CC molecules which correspond to the SESTs. Proteins encoded by the SESTs  
 CC are useful in assays for determining biological activity and raising  
 CC antibodies. They may be useful for treatment of autoimmune disorders  
 CC (multiple sclerosis, insulin dependent diabetes), allergic conditions  
 CC (asthma), myeloid or lymphoid cell deficiencies, wounds, burns, ulcers,  
 CC osteoporosis, osteoarthritis, central nervous system disorders  
 CC (Alzheimer's, Parkinson's, Huntington's, inflammatory disorders (Crohn's  
 CC disease), tumours, bacterial, fungal or viral infections, depression and  
 CC psoriasis. AAA45926 to AAA45931 represent linker variants which are given  
 CC in the exemplification of the present invention.  
 XX  
 SQ Sequence 304 BP; 81 A; 67 C; 75 G; 81 T; 0 other;  
 XX  
 Query Match 5.6%; Score 18; DB 21; Length 304;  
 Best Local Similarity 100.0%; Pred. No. 10;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 XX  
 QY 105 CAACACTTTTCATGTCCT 122  
 DB |||||  
 DB 171 caaacactttcatgtcct 188  
 XX  
 RESULT 11  
 AAH50777  
 ID AAH50777 standard; cDNA; 332 BP.  
 XX  
 AC AAH50777;  
 XX  
 XX 23-AUG-2001 (first entry)  
 DT Human tumour associated cDNA #106.  
 DE  
 XX Human; cancer specific gene expression; gene therapy;  
 KW age related differential expression; ss.  
 KW  
 XX Homo sapiens.  
 OS

```
XX PN W0200136685-A2.
XX PD
XX PF 25-MAY-2001.
XX PR 17-NOV-2000; 2000WO-US31809.
XX PR 17-NOV-1999; 99US-0166056.
XX PR 17-NOV-1999; 99US-0166106.
XX PA (NYXI-) NYXIS NEURO THERAPIES INC.
XX PI Kroes RA, Moskal JR, Yamamoto H;
XX DR WPI; 2001-355647/37.
XX PT Novel nucleic acid molecules differentially expressed in brain cancers,
XX PT useful for ascertaining propensity of cell for malignant phenotype or
XX PT ascertaining suitability of anti-neoplastic drug candidate -
XX PS Claim 28; Page 52; 82pp; English.
XX CC The present invention provides the sequences of 184 cDNA fragments which
XX CC are differentially expressed in cancer cell depending on the age of the
XX CC patient. They can be used to diagnose and identify treatments for
XX CC cancers, particularly brain cancers such as haemangioblastoma, teratoma,
XX CC haemangioma, glioblastoma, schwannoma, osteoma and pinealoma. The
XX CC present sequence is a cancer-associated cDNA of the invention.
XX SQ Sequence 332 BP; 95 A; 45 C; 72 G; 120 T; 0 other;

Query Match 5.6%; Score 18; DB 22; Length 332;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 168 TCAATTTGTGAAGATC 185
Db 79 tcaattgtgaagatc 96

RESULT 12
AAX02011
ID AAX02011 standard; DNA; 1398 BP.
XX AC AAX02011;
XX DT 21-APR-1999 (first entry)
XX DE Human HDRG DNA.
XX KW Human developmentally regulated GTP-binding protein; HDRG; cancer;
XX KW diagnosis; treatment; prevention; immune disorder; neoplastic disorder;
XX KW smooth muscle disorder; ss.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX FT CDS 64..1167
XX FT /*tag= a
XX FT /product= "HDRG"
XX PN US5871971-A.
XX PD 16-FEB-1999.
XX PF 17-JUL-1997; 97US-0896320.
XX PR 17-JUL-1997; 97US-0896320.
XX PA (INCY-) INCYTE PHARM INC.
XX PI Corley NC, Hillman JL, Lal P, Shah P;
```

```
XX DR WPI; 1999-166645/14.
XX DR P-ESDB; AAW92455.
XX PT New polynucleotides encoding human developmentally regulated
XX PT GTP-binding protein (HDRG) - useful for diagnosing, preventing and
XX PT treating immune, abnormal vesicle trafficking, smooth muscle and
XX PT neoplastic disorders
XX PS Claim 1; Fig 1A-D; 32pp; English.
XX CC This sequence encodes a human developmentally regulated GTP-binding
XX CC protein (HDRG). The HDRG protein can be used as a probe to detect a
XX CC polynucleotide which encodes HDRG. The polynucleotide sequences encoding
XX CC HDRG (or complements) may be used for the diagnosis of conditions or
XX CC disorders associated with expression of HDRG and treatment/prevention
XX CC of immune, abnormal vesicle trafficking, smooth muscle and neoplastic
XX CC disorders e.g. AIDS, Addison's disease, atherosclerosis, bronchitis,
XX CC ulcerative colitis, diabetes mellitus, emphysema, gout, Graves' disease,
XX CC osteoporosis, rheumatoid arthritis, Sjogren's syndrome and
XX CC complications of cancer. The polynucleotides may also be used in Southern
XX CC or Northern analysis, dot blot, or other membrane based technologies, in
XX CC PCR technologies or in dipstick, pin, ELISA assays or microarrays
XX CC utilizing fluids or tissues from patient biopsies to detect altered
XX CC HDRG expression.
XX SQ Sequence 1398 BP; 394 A; 302 C; 354 G; 348 T; 0 other;

Query Match 5.6%; Score 18; DB 20; Length 1398;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 167 TTCAAAATTGTGAAGAAGT 184
Db 1148 ttcaaatgtgaagaagt 1165

RESULT 13
AA154116
ID AA154116 standard; DNA; 182 BP.
XX AC AA154116;
XX DT 17-OCT-2001 (first entry)
XX DE Probe #22802 used to measure gene expression in human placenta sample.
XX KW Probe; microarray; human; placenta; antenatal diagnosis;
XX KW genetic disorder; ss.
XX OS Homo sapiens.
XX PN WQ200157272-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US00663.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX DR WFI; 2001-488897/53.
```

PT Human genome-derived single exon nucleic acid probes useful for  
 PT analyzing gene expression in human placenta -  
 XX  
 PS Claim 25; SEQ ID No 22802; 654pp; English.  
 XX  
 CC The present invention relates to single exon nucleic acid probes (SENP).  
 CC The present sequence is one such probe. The probes are useful for  
 CC producing a microarray for predicting, measuring and displaying gene  
 CC expression in samples derived from human placenta. The probes are useful  
 CC for antenatal diagnosis of human genetic disorders.  
 XX  
 SQ Sequence 182 BP; 50 A; 36 C; 58 G; 38 T; 0 other;

Query Match 5.3%; Score 17; DB 22; Length 182;  
 Best Local Similarity 100.0%; Pred. No. 33;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 166 TTTCAAATTCGAGAA 182  
 |||||  
 DB 60 ttcaaatgtgaagaa 76

RESULT 14  
 AAF68447  
 ID AAF68447 standard; cDNA; 528 BP.  
 XX  
 AC AAF68447;  
 XX  
 DT 12-APR-2001 (first entry)  
 XX  
 DE Human lung tumour protein related nucleotide sequence SEQ ID NO:382.  
 XX  
 KW Human; lung cancer; lung tumour; lung tumour protein; gene therapy;  
 KW lung cancer antigen; lung tumour-specific antigen; diagnosis; vaccine;  
 KW cytostatic; antisense inhibition; ss.  
 XX  
 OS Homo sapiens.

XX WO200100828-A2.  
 PN 04-JAN-2001.  
 XX  
 PD 30-JUN-2000; 2000WO-US18061.  
 PF  
 XX 30-JUN-1999; 99US-0346492.  
 PR 15-OCT-1999; 99US-0419356.  
 PR 17-DEC-1999; 99US-0466867.  
 PR 30-DEC-1999; 99US-0476300.  
 PR 08-MAR-2000; 2000US-0519642.  
 PR 22-MAR-2000; 2000US-0533077.  
 PR 10-APR-2000; 2000US-0546259.  
 PR 27-APR-2000; 2000US-0560406.  
 PR 05-JUN-2000; 2000US-0589184.  
 XX  
 PA (CORI-) CORIXA CORP.  
 XX  
 XX Wang T, Bangur CS, Lodes MJ, Fanger GR, Vedvick TS, Carter D;  
 PI Retter MW, Mannion J;  
 XX WPI; 2001-071488/08.  
 DR  
 XX Lung tumor-associated proteins and the nucleic acids that encode them,  
 PT useful for preventing, diagnosing and treating lung cancer -  
 PT  
 XX Claim 4; Page 285; 436pp; English.  
 XX  
 CC The present invention describes immunogenic portions of lung tumour-  
 CC associated proteins (I) and the nucleic acids (NAs) that encode them.  
 CC (I) have cytostatic activity and can be used in gene therapy, antisense  
 CC inhibition and in vaccines. The NAs and the lung tumour-associated  
 CC proteins they encode may be used in the prevention, treatment and  
 CC diagnosis of diseases associated with their inappropriate expression,

CC especially lung cancers. For example, the NAs may be administered to  
 CC treat diseases by rectifying mutations or deletions in a patient's genome  
 CC that affect the activity of the protein by expressing inactive proteins  
 CC or to supplement the patients own production of (I). Additionally, the  
 CC NAs may be used to produce the lung-tumour associated protein, according  
 CC to standard recombinant DNA methodology. Conversely, antisense NA  
 CC molecules may be administered to down regulate protein expression by  
 CC binding with the cells own genes and preventing their expression. The NA  
 CC and complementary sequences may also be used as DNA probes in diagnostic  
 CC assays to detect and quantitate the presence of similar NA sequences in  
 CC samples, and hence which patients may be in need of treatment for lung  
 CC cancer. The (I) may be used as antigens in the production of antibodies  
 CC and in assays to identify modulators (agonists and antagonists) of the  
 CC expression and activity of the protein. AAF68083 to AAF68878 and  
 CC AAF6848 to AAF6878 represent human lung tumour protein related  
 CC nucleotide and protein sequences which are used in the exemplification  
 CC of the present invention.

XX Sequence 528 BP; 112 A; 141 C; 143 G; 130 T; 2 other;

Query: Match 5.3%; Score 17; DB 22; Length 528;  
 Best Local Similarity 100.0%; Pred. No. 34;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 250 GCCTCAGTAGATTTT 266  
 |||||  
 DB 330 gcctcagtagattttt 346

RESULT 15  
 AAI41051  
 ID AAI41051 standard; DNA; 538 BP.  
 XX  
 AC AAI41051;  
 XX  
 DT 17-OCT-2001 (first entry)  
 XX  
 DE Probe #9737 used to measure gene expression in human placenta sample.  
 XX Probe; microarray; human; placenta; antenatal diagnosis;  
 KW genetic disorder; ss.  
 XX  
 OS Homo sapiens.

XX WO200157272-A2.  
 XX C9-AUG-2001.  
 XX 30-JAN-2001; 2001WO-US00663.  
 XX 04-FEB-2000; 2000US-0180312.  
 PR 26-MAY-2000; 2000US-0207456.  
 PR 30-JUN-2000; 2000US-0608408.  
 PR 03-AUG-2000; 2000US-0632366.  
 PR 21-SEP-2000; 2000US-0234687.  
 PR 27-SEP-2000; 2000US-0236359.  
 PR 04-OCT-2000; 2000GB-0024263.  
 XX  
 PA (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;  
 PI WPI; 2001-48897/53.  
 XX  
 XX Human genome-derived single exon nucleic acid probes useful for  
 PT analyzing gene expression in human placenta -  
 PT  
 XX Claim 25; SEQ ID No 9737; 654pp; English.

XX The present invention relates to single exon nucleic acid probes (SENP).  
 CC The present sequence is one such probe. The probes are useful for  
 CC producing a microarray for predicting, measuring and displaying gene



CC expression in samples derived from human placenta. The probes are useful  
CC for antenatal diagnosis of human genetic disorders.

XX  
SQ Sequence 538 BP; 159 A; 104 C; 123 G; 152 T; 0 other;

Query Match 5.3%; Score 17; DB 22; Length 538;  
Best Local Similarity 100.0%; Pred. No. 34;  
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 166 TTTCAAATTGTGAGAA 182

|||||

Db 378 ttccaatttgagaaa 394

Search completed: March 30, 2002, 07:13:51  
Job time: 3978 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 30, 2002, 08:13:48 ; Search time 119.38 Seconds  
(without alignments)  
149.483 Million cell updates/sec

Title: US-09-095-478A-8  
Perfect score: 673  
Sequence: 1 DFGGMWNNCNIAMITRE.....VRKSHITGLLVHCTAGVR 122

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

- Database : SPTREMBL.17.\*
- 1: sp.archaea.\*
  - 2: sp.bacteria.\*
  - 3: sp.fungi.\*
  - 4: sp.human.\*
  - 5: sp.invertebrate.\*
  - 6: sp.mammal.\*
  - 7: sp.mbc.\*
  - 8: sp.organelle.\*
  - 9: sp.phage.\*
  - 10: sp.plant.\*
  - 11: sp.rodent.\*
  - 12: sp.virus.\*
  - 13: sp.vertebrate.\*
  - 14: sp.unclassified.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	566	84.1	426	11	O55082	O55082 mus musculus
2	529	78.6	398	4	Q9Y406	Q9Y406 homo sapien
3	284.5	42.3	2484	6	Q28006	Q28006 bos taurus
4	274.5	40.8	2460	11	Q64512	Q64512 mus musculus
5	261	38.8	1156	5	Q21214	Q21214 caenorhabdi
6	260.5	38.7	121	4	Q9UDA8	Q9uda8 homo sapien
7	255.5	38.0	1216	13	Q9IBD8	Q9ibd8 cyprinus ca
8	253.5	37.7	1245	13	Q9IBF0	Q9ibf0 fugu rubrip
9	253.5	37.7	1246	13	Q9IBF1	Q9ibf1 fugu rubrip
10	243.5	36.2	578	5	Q9W324	Q9w324 drosophila
11	240.5	35.7	597	13	Q9IB71	Q9ib71 xenopus lae
12	240	35.7	579	11	Q9JJ07	Q9jj07 mus musculu
13	239	35.5	1337	4	Q15255	Q15255 homo sapien
14	238	35.4	832	13	Q90947	Q90947 gallus gall
15	238	35.4	1648	4	Q9UDR6	Q9udr6 homo sapien
16	237	35.2	468	13	Q9IBA0	Q9iba0 potamotrygo
17	237	35.2	469	13	Q9NL08	Q9nl08 eptatretus
18	237	35.2	615	13	Q9IAI8	Q9ia18 xenopus lae
19	237	35.2	1399	4	O75870	O75870 homo sapien

20	237	35.2	1499	13	Q90815	Q90815 gallus gall
21	237	35.2	1502	4	Q9UM81	Q9um81 homo sapien
22	237	35.2	1948	4	Q13332	Q13332 homo sapien
23	235.5	35.0	1291	11	Q61812	Q61812 mus musculu
24	235	34.9	468	13	Q9NL06	Q9nl06 eptatretus
25	233.5	34.7	694	13	Q91870	Q91870 xenopus lae
26	233	34.6	1216	11	Q82884	Q82884 rattus norv
27	232.5	34.5	811	4	Q9UI21	Q9ui21 homo sapien
28	232.5	34.5	813	4	Q9UI22	Q9ui22 homo sapien
29	232.5	34.5	823	4	Q9Y4V5	Q9y4v5 homo sapien
30	232.5	34.5	1436	4	Q92850	Q92850 homo sapien
31	232.5	34.5	1436	4	O00197	O00197 homo sapien
32	232.5	34.5	1439	4	P78399	P78399 homo sapien
33	232	34.5	361	11	Q61373	Q61373 mus musculu
34	232	34.5	468	13	Q91BA2	Q91ba2 potamotrygo
35	232	34.5	1896	13	Q9IAJ1	Q9iaj1 xenopus lae
36	231	34.3	1501	11	Q9QW00	Q9qw00 rattus sp.
37	231	34.3	1863	11	O64605	O64605 rattus norv
38	231	34.3	1904	11	O64699	O64699 mus musculu
39	230.5	34.2	1114	4	Q9H0Y6	Q9hy06 homo sapien
40	230.5	34.2	1143	4	O16614	O16614 homo sapien
41	230	34.2	460	11	O62917	O62917 rattus norv
42	230	34.2	582	11	O64696	O64696 mus musculu
43	230	34.2	1406	13	Q9W6V5	Q9w6v5 gallus gall
44	230	34.2	1887	11	Q9QW67	Q9qw67 rattus sp.
45	230	34.2	1898	11	O64604	O64604 r protein-t

ALIGNMENTS

RESULT 1  
Q55082  
ID O55082 PRELIMINARY: PRT: 426 AA.  
AC O55082;  
DT 01-JUN-1998 (TREMBLrel. 06, Created)  
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)  
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)  
DE PROTEIN TYROSINE PHOSPHATASE, NON-RECEPTOR TYPE 20 (EC 3.1.3.48)  
DE (PHOSPHOTYROSINE PHOSPHATASE) (PTPASE).  
GN PTPN20.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
CC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Mus.  
CX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=TESTIS;  
RX MEDLINE=98070510; PubMed=9407093;  
RA Ohnogi M., Kuramochi S., Matsuda S., Yamamoto T.;  
RT "Molecular cloning and characterization of a novel cytoplasmic  
RT n-protein-tyrosine phosphatase that is specifically expressed in  
RT spermatocytes.";  
RL J. Biol. Chem. 272:33092-33099(1997).  
GC 1- CATALYTIC ACTIVITY: PROTEIN TYROSINE PHOSPHATE + H(2)O -> PROTEIN  
CC TYROSINE + PHOSPHATE.  
DR EMBL; D64141; BAA23761.1; -  
DR HSSP; P28827; IRPM.  
DR MGD; MGI:1196295; Ptpn20.  
DR InterPro; IPR000387; TYR\_phosphatase.  
DR InterPro; IPR000242; Tyr\_prot\_phptase.  
DR Pfam; PF00102; Y\_phosphatase; 1.  
DR PRINTS; PR00700; PRTYPHPTASE.  
DR SMART; SM00194; PTPC; 1.  
DR PROSITE; PS00383; TYR\_PHOSPHATASE\_1; 1.  
DR PROSITE; PS00056; TYR\_PHOSPHATASE\_2; 1.  
DR PROSITE; PS00055; TYR\_PHOSPHATASE\_PTP; 1.  
KW Hydrolyase.  
SQ SEQUENCE 426 AA; 49118 MW; 2B35FBI3379502F4 CRC64;

Der 26

Query/Match 84.1%; Score 566; DB 11; Length 426;  
Best Local Similarity 83.6%; Pred. No. 1.1e-55;

```

OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
SEQUENCE FROM N.A.
RP Vega Q.C., Walton K.M., Dixon J.E.;
RA Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
RL -1: CATALYTIC ACTIVITY: PROTEIN TYROSINE PHOSPHATE + H(2)O -> PROTEIN
CC TYROSINE + PHOSPHATE.
CC -1: SIMILARITY: CONTAINS A DOMAIN FOUND IN BAND 4.1, EZRIN, MOESIN,
CC RADIXIN, AND TALIN.
CC -1: SIMILARITY: BELONGS TO THE NON-RECEPTOR CLASS OF THE PROTEIN-
CC TYROSINE PHOSPHATASE FAMILY.
CC EMBL: U20807; AAA73516.1;
DR HSSP: Q12923; 3PDZ.
DR InterPro: IPR000299; Band_4.1.
DR InterPro: IPR001478; PDZ.
DR InterPro: IPR000387; TYR_phosphatase.
DR InterPro: IPR000242; TYR_prot_phptase.
DR Pfam: PF00373; Band_41; 1.
DR Pfam: PF00595; PDZ; 5.
DR Pfam: PF00102; Y_phosphatase; 1.
DR PRINTS: PR00935; BAND41.
DR PRINTS: PR00700; PRTYPHPHTASE.
DR SMART: SM00295; B41; 1.
DR SMART: SM00228; PDZ; 5.
DR SMART: SM00194; FTPC; 1.
DR PROSITE: PS00660; BAND_41_1; FALSE_NEG.
DR PROSITE: PS00661; BAND_41_2; FALSE_NEG.
DR PROSITE: PS00557; BAND_41_3; 1.
DR PROSITE: PS0106; PDZ; 5.
DR PROSITE: PS00383; TYR_PHOSPHATASE_1; FALSE_NEG.
DR PROSITE: PS00556; TYR_PHOSPHATASE_2; 1.
DR PROSITE: PS00555; TYR_PHOSPHATASE_PTP; 1.
DR KW: Structural protein; Cytoskeleton; Hydrolase; Coiled coil.
FT DOMAIN 600 800
FT FT DOMAIN 1341 1344 POLY-SER.
FT FT DOMAIN 2236 2484 PROTEIN-TYROSINE PHOSPHATASE.
FT ACT_SITE 2407 2407 BY SIMILARITY.
FT FT DOMAIN 467 504 COILED COIL (POTENTIAL).
FT FT DOMAIN 1761 1793 COILED COIL (POTENTIAL).
SQ SEQUENCE 2484 AA; 276382 MW; 45A92F0D4F1ED13D CRC64;

Query Match 42.3%; Score 284.5; DB 6; Length 2484;
Best Local Similarity 42.3%; Pred. No. 4.6e-23;
Matches 52; Conservative 21; Mismatches 49; Indels 1; Gaps

QY 1 DFGWMWENNCNVIAMITIEGGVIKCCSWPVSL-KEPFEKHFHVHLENFOITQYVF 59
DB 2291 DFQMWIWEQNSVIAAMTQVEGEKIKORYWPNVLGKSTWVSNRLRLALVRVQOLKGFV 2350
QY 60 IRIFQIVKSTGKSHSVKHLKIPKPDHCTPQASVDFFIKYVYRKSHITGLPLLVHCTAG 119
DB 2351 VRAMTLEDIQTGEVRHVSHLNTAFPDHDTQPQDDLTFTSYMRHVSRSPIITHCSAG 2410
QY 120 VGR 122
DB 2411 IGR 2413

RESULT 4
ID Q64512 PRELIMINARY; PRT: 2460 AA.
AC Q64512; Q62135; Q61494; Q64499;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE PROTEIN-TYROSINE PHOSPHATASE, NONRECEPTOR-TYPE.13 (EC 3.1.3.48)
DE (PROTEIN-TYROSINE PHOSPHATASE RIP) (PHOSPHOPROTEIN PHOSPHATASE)
DE (PROTEIN-TYROSINE-PHOSPHATASE) (PHOSPHOTYROSINE PHOSPHATASE) (PTPASE)

```





Db 642 DFWRMVWEEKSSIIYVMTCEEGNKKCAQYWPDLREAEIEEFVVKIRSEBHCDDYVI 701  
 QY 61 R-IFQIVKSTGKSHVKHLQIKWPDHGTPASVDFFIKYRVYR--KSHITGPLLHVHCT 117  
 Db 702 RHLLLNKKREKGSREVTHIQISWPDHGVDPGDSLLKLRNVNSFKNFSGPVVHCS 761  
 QY 118 AGVGR 122  
 Db 762 AGVGR 766

RESULT 8  
 Q918F0 PRELIMINARY; PRT; 1245 AA.  
 ID Q918F0;  
 AC Q918F0;  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
 DE CD45 (EC 3.1.3.48).  
 GN PTPRC.  
 OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;  
 OC Tetraodontidae; Takifugu.  
 OX NCBI\_TaxID=31033;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=SPLEEN;  
 RA Diaz del Pozo E.M., Beverley P.C., Timon M.;  
 RT "Genomic structure and sequence of the leukocyte common antigen (CD45) from the pufferfish Fugu rubripes and comparison with its mammalian homologue."  
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AJ243430; CAB96212.1;  
 DR InterPro; IPR003961; FN.III.  
 DR InterPro; IPR000387; TYR\_phosphatase.  
 DR Pfam; PF00041; fn3; 1.  
 DR PRINTS; PR00700; PRTYPHPHTASE.  
 DR SMART; SM00194; PTPC; 2.  
 DR PROSITE; PS00383; TYR\_PHOSPHATASE\_1; 2.  
 DR PROSITE; PS50056; TYR\_PHOSPHATASE\_2; 2.  
 DR PROSITE; PS50055; TYR\_PHOSPHATASE\_PTP; 2.  
 DR Hydrolase; Signal.  
 FT SIGNAL 1 19 POTENTIAL.  
 PT CHAIN 20 1246 CD45.  
 SQ SEQUENCE 1245 AA; 141324 MW; 6CB711EFF5797555 CRC64;

Query Match 37.7%; Score 253.5; DB 13; Length 1245;  
 Best Local Similarity 39.2%; Pred. No. 6.5e-20;  
 Matches 49; Conservative 21; Mismatches 52; Indels 3; Gaps 2;  
 QY 1 DFWGMWENNVCNVIAMITREIEGGVKKCSYWPVSLKEPLEKHFHVLLENFOITQYFVI 60  
 Db 700 NFWRMVWEEQTSIIYVNSRCEGNRIKCAQYWPSEDRTEIEEFIVKLTSEDHYPDYII 759  
 QY 61 RIFQIV-KKSTGKSHVKHLQIKWPDHGTPASVDFFIKYRVYR--KSHITGPLLHVHCT 117  
 Db 760 RHLSLTNKKDKGSREVTHIQISWPDHGVPEEAQALLKLRNVNSFKNFSGPVVHCS 819  
 QY 118 AGVGR 122  
 Db 820 AGVGR 824

RESULT 9  
 Q918F1 PRELIMINARY; PRT; 1246 AA.  
 ID Q918F1;  
 AC Q918F1;  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)

DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
 DE CD45-PRECURSOR (EC 3.1.1.3.48).  
 GN PTPRC.  
 OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;  
 OC Tetraodontidae; Takifugu.  
 OX NCBI\_TaxID=31033;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=SPLEEN;  
 RA Diaz del Pozo E.M., Beverley P.C., Timon M.;  
 RT "Genomic structure and sequence of the leukocyte common antigen (CD45) from the pufferfish Fugu rubripes and comparison with its mammalian homologue."  
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AJ243429; CAB96211.1;  
 DR InterPro; IPR003961; FN.III.  
 DR InterPro; IPR000387; TYR\_phosphatase.  
 DR Pfam; PF00041; fn3; 1.  
 DR PRINTS; PR00700; PRTYPHPHTASE.  
 DR SMART; SM00060; FN3; 1.  
 DR SMART; SM00194; PTPC; 2.  
 DR PROSITE; PS00383; TYR\_PHOSPHATASE\_1; 2.  
 DR PROSITE; PS50056; TYR\_PHOSPHATASE\_2; 2.  
 DR PROSITE; PS50055; TYR\_PHOSPHATASE\_PTP; 2.  
 DR Hydrolase; Signal.  
 FT SIGNAL 1 19 POTENTIAL.  
 PT CHAIN 20 1246 CD45.  
 SQ SEQUENCE 1246 AA; 141363 MW; 4643259F5CA48E8E CRC64;

i:Query:Match 37.7%; Score 253.5; DB 13; Length 1246;  
 i:Best Local Similarity 39.2%; Pred. No. 6.5e-20;  
 i:Matches 49; Conservative 21; Mismatches 52; Indels 3; Gaps 2;  
 QY 1 DFWGMWENNVCNVIAMITREIEGGVKKCSYWPVSLKEPLEKHFHVLLENFOITQYFVI 60  
 Db 700 NFWRMVWEEQTSIIYVNSRCEGNRIKCAQYWPSEDRTEIEEFIVKLTSEDHYPDYII 760  
 QY 61 RIFQIV-KKSTGKSHVKHLQIKWPDHGTPASVDFFIKYRVYR--KSHITGPLLHVHCT 117  
 Db 760 RHLSLTNKKDKGSREVTHIQISWPDHGVPEEAQALLKLRNVNSFKNFSGPVVHCS 820  
 QY 118 AGVGR 122  
 Db 820 AGVGR 825  
 RESULT 10  
 Q9W324 PRELIMINARY; PRT; 578 AA.  
 ID Q9W324;  
 AC Q9W324;  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)  
 DE CG3101 PROTEIN.  
 GN CG3101.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,

AC	Q9I87L;
DT	01-NOV-1996 (TrEMBLrel. 01, Created)
DT	01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT	01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE	Tyrosine phosphatase [EC 3.1.3.48].
OS	Xenopus laevis (African clawed frog).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
OC	Xenopodinae; Xenopus.
OX	NCBI_TaxID=8355;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	TISSUE=OVARY;
RX	MEDLINE=94308257; PubMed=8034733;
RA	Del Vecchio R.L., Tonks N.K.;
RT	"Characterization of two structurally related Xenopus laevis protein
RT	tyrosine phosphatases with homology to lipid-binding proteins.";
NL	J. Biol. Chem. 269:19639-19645(1994).
DR	EMBL; L33099; AAA21728.1; -
DR	HSP; P18052; IYFO.
DR	InterPro: IPR001251; CRAL_TRIO.
DR	InterPro: IPR000387; TYR_phosphatase.
DR	InterPro: IPR000242; Tyr_prot_phptase.
DR	Pfam; PF00650; CRAL_TRIO; 1.
DR	Pfam; PF00102; Y_phosphatase; 1.
DR	PRINTS; PR00700; PTYPHPPTASE.
DR	SMART; SM00194; PTPc; 1.
DR	SMART; SM00516; SEC14; 1.
DR	PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
DR	PROSITE; PS00056; TYR_PHOSPHATASE_2; 1.
DR	PROSITE; PSS00055; TYR_PHOSPHATASE_PTP; 1.
KW	Hydrolase.
SQ	SEQUENCE 597 AA: 68842 MW: 68429DEF718B0735 CRC64;
	Query Match 35.7%; Score 240.5; DB 13; Length 597;
	Best local Similarity 35.0%; Pred. No. 8.3e-19;
	Matches 48; Conservative 21; Mismatches 53; Indels 15; Gaps
CY	1 DFWGMHNNCNVIAMITREIEGVIVKCCSYWPYSLKPEFKFHVLENFQITQYFI 60
	:    :            :    :    :    :
DB	366 DFNRMVVEQVLIIVMTRVIERGRICGOWPLEAGRSDETGTFLIRNIHIDLFDQDKL 44
	:    :            :    :    :    :
QY	61 RIFOIVKSTCKTSKSHVKHLQFIKPDDHGTPASVDFFTKIVRYVRK-----S 10
	:    :    :    :            :    :    :    :
DB	446 THEEVYNKKQDDESRAHYQTMSPDGVPKSKASAMILDQRSQKHQAQAVQNLCMGWTG 50

RESULT	12
Q9JY07	
ID	PRELIMINARY; PRT; 579 AA.
AC	Q9JJ07;
ET	01-OCT-2000 (TEMBLrel. 15, Created)
DT	01-OCT-2000 (TEMBLrel. 15, Last sequence update)
DT	01-JUN-2001 (TEMBLrel. 17, Last annotation update)
DE	RECEPTOR-TYPE PROTEIN TYROSINE PHOSPHATASE [FRAGMENT].
CS	Mus musculus (Mouse).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;	
[1]	SEQUENCE FROM N.A.
RN	TISSUE-VASCULAR ENDOTHELIIUM;
RA	Fachinger G., Deutsch U., Risau W.;
RT	"Functional interaction of vascular endothelial protein tyrosine
RL	phosphatase with the angiopeptin receptor tie-2.";
RI	Oncogene 0:0-0(1999).
DR	EMBL; AF157628; AAF80346.1; -.
DR	InterPro: IPR000387; Tyr.phosphatase.



```

DR InterPro: IPR000242; Tyr_prot_phptase.
DR Pfam: PF00102; Y_phosphatase; 1.
DR PRINTS: PR00700; PTYPHPTASE.
DR SMART: SM00194; PTPC; 1.
DR PROSITE: PS00383; TYR_PHOSPHATASE_1; 1.
DR PROSITE: PS00056; TYR_PHOSPHATASE_2; 1.
DR PROSITE: PS00055; TYR_PHOSPHATASE_PTP; 1.
KW Hydrolase.
FT NON_TER 1
SQ SEQUENCE 579 AA; 66298 MW; 67475838046C870 CRC64;

Query Match 35.7%; Score 240; DB 11; Length 579;
Best Local Similarity 36.2%; Pred. No. 9.1e-19;
Matches 46; Conservative 26; Mismatches 49; Indels 6; Gaps 3;

Qy 1 DFWGMWNNCNVIAITREIEGVKCCSWPVSLEKPELFKHFHVLLENFQITQYFVI 60
    ||| ||| : : : : : ||| ||| : : : : : : : : : : : : : : : : : :
Dy 367 DFWKMAEQNVHNVITVQCVEKGRVKCDHYWPAQDPLYYGDLILQMVSESLPEWTI 425
    ||| ||| : : : : : ||| ||| : : : : : : : : : : : : : : : : : :
Qy 61 RIFQIVKSTGKSHS-VKHLOFIKWPDHGTTPASVDFFIKYR----YVRKSHITGPLLVH 115
    ||| : : : : : ||| ||| : : : : : ||| : : : : : ||| : : : : :
Dy 426 REFKICSEQLDAHLIRHFHYTVPDGHGVPETTSQSLIQFVRTVRDYNRSPGAGPTVVH 485
    ||| ||| : : : : : ||| ||| : : : : : ||| : : : : : ||| : : : : :
Qy 116 CTAGVGR 122
    ||| ||| |||
Dy 486 CSAGVGR 492
    ||| ||| |||

RESULT 13
ID Q15255 PRELIMINARY; PRT; 1337 AA.
AC Q15255..
DT 01-NOV-1996 (TremBLrel. 01, Created)
DT 01-NOV-1996 (TremBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TremBLrel. 17, Last annotation update)
DE PROTEIN-TYROSINE PHOSPHATASE ETA PRECURSOR (EC 3.1.3.48) (R-PTP-ETA).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95086212; PubMed=7994032;
RA Honda H., Inazawa J., Nishida J., Yazaki Y., Hirai H.;
RT "Molecular cloning, characterization, and chromosomal localization of
RT a novel protein-tyrosine phosphatase, HPTP eta.";
RL Blood 84:4186-4194(1994).
DR EMBL: D37781; BAA07035.1; .
DR HSP: P18052; 1YFO.
DR InterPro: IPR001395; Aldo_ket_red.
DR InterPro: IPR003961; FN_III.
DR InterPro: IPR000387; Tyr_phosphatase.
DR InterPro: IPR000242; Tyr_prot_phptase.
DR Pfam: PF00041; fn3; 5.
DR Pfam: PF00102; Y_phosphatase; 1.
DR PRINTS: PR00700; PTYPHPTASE.
DR SMART: SM00060; FN3; 6.
DR SMART: SM00194; PTPC; 1.
DR PROSITE: PS00063; ALDOXETO_REDUCTASE_3; UNKNOWN_1.
DR PROSITE: PS00383; TYR_PHOSPHATASE_1; 1.
DR PROSITE: PS00056; TYR_PHOSPHATASE_2; 1.
DR PROSITE: PS00055; TYR_PHOSPHATASE_PTP; 1.
KW Glycoprotein; Hydrolase; Repeat; Signal; Transmembrane.
FT SIGNAL 1 35 POTENTIAL.
FT CHAIN 36 1337 PROTEIN-TYROSINE PHOSPHATASE ETA.
FT DOMAIN ? 1337 FIBRONECTIN TYPE-III.
FT DOMAIN ? 1337 PROTEIN-TYROSINE PHOSPHATASE.
FT TRANSMEM 36 975 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 976 996 POTENTIAL.
FT DOMAIN 997 1337 CYTOPLASMIC (POTENTIAL).
FT ACT_SITE 1239 1239 BY SIMILARITY.
FT CARBOHYD 72 72 N-LINKED (GLCNAC. .) (POTENTIAL).

```

```

FT CARBOHYD 82 82 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 93 93 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 104 104 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 142 142 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 172 172 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 192 192 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 231 231 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 258 258 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 278 278 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 342 342 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 351 351 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 376 376 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 391 391 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 396 396 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 413 413 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 431 431 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 501 501 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 525 525 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 536 536 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 582 582 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 603 603 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 618 618 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 628 628 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 637 637 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 666 666 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 669 669 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 761 761 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 772 772 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 784 784 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 790 790 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 824 824 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 910 910 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 937 937 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 1337 AA; 145801 MW; BA7EA3EE32961CEA CRC64;

```

```

Query Match 35.5%; Score 239; DB 4; Length 1337;
Best Local Similarity 38.1%; Pred. No. 3.1e-18;
Matches 48; Conservative 22; Mismatches 50; Indels 6; Gaps 3;

Qy 1 DFWGMWNNCNVIAITREIEGVKCCSWPVSLEKPELFKHFHVLLENFQITQYFVI 60
    ||| ||| : : : : : ||| ||| : : : : : : : : : : : : : : : : : :
Dy 1122 DFWKMAEQNVHNVITVQCVEKGRVKCDHYWPAQDPLYYGDLILQMVSESLPEWTI 1179
    ||| ||| : : : : : ||| ||| : : : : : ||| : : : : : ||| : : : : :
Qy 61 RIFQIVKSTGKSHS-VKHLOFIKWPDHGTTPASVDFFIKYR---YVRKSHITGPLLVH 116
    ||| : : : : : ||| ||| : : : : : ||| : : : : : ||| : : : : :
Dy 1180 RDTVKNIQTSSEHPLRQFHFTSWPDGHGVPDITDILLINFRYLVRDYMVKQSPESILVHC 1239
    ||| ||| : : : : : ||| ||| : : : : : ||| : : : : : ||| : : : : :
Qy 117 TAGVGR 122
    ||| ||| |||
Dy 1240 SAGVGR 1245
    ||| ||| |||

RESULT 14
ID Q90947 PRELIMINARY; PRT; 832 AA.
AC Q90947.
DT 01-NOV-1996 (TremBLrel. 01, Created)
DT 01-NOV-1996 (TremBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TremBLrel. 17, Last annotation update)
DE PHOSPHOTYROSYL PHOSPHATASE (EC 3.1.3.48).
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=941122757; PubMed=8293038;
RA Rowley R., Lee J.M., Corbell H.B., Charbonneau R., Jue K.,
RA Dankort D.L., Branton P.E.;
RT "Isolation of chicken phosphotyrosyl phosphatase cDNA sequences and

```

RT identification of a brain-specific species related to human PTP

```
RT zeta.";
RL Cell. Mol. Biol. Res. 39:209-219(1993).
DR EMBL; L27625; AAA49015.1; -.
DR HSP; P18052; IYFO.
DR InterPro; IPR000387; Tyr_phosphatase.
DR InterPro; IPR000242; Tyr_prot_phptase.
DR Pfam; PF00102; Y_phosphatase; 2.
DR PRINTS; PR00700; PRTYPHPTASE.
DR SMART; SM00194; PTPC; 2.
DR PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
DR PROSITE; PS00056; TYR_PHOSPHATASE_2; 2.
DR PROSITE; PS00055; TYR_PHOSPHATASE_PTP; 2.
KW Hydrolase.
SQ SEQUENCE 832 AA; 93297 MW; 4A61F365B8D794F3 CRC64;

Query Match 35.4%; Score 238; DB 13; Length 832;
Best Local Similarity 37.0%; Pred. No. 2.3e-18;
Matches 50; Conservative 24; Mismatches 43; Indels 18; Gaps 5;

QY 1 DFWGMMNENNCVIAITREIEGGVIKCCSYWPVSLKEPLEFKHFLHLENFQITQYFVI 60
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 327 DFWRMIEHNVEVIVMITNLLEKGRKCDQYWPAGSE--EYGNFLVTQKSVHVLAYTV 384
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 61 RIFOI----VKKSTGKSHS-----VKHLQFIKWPDHGTPTASVDFFIKYVRYVRKS-----H 107
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 385 RNFTLRNTKIKKSGKGRVVTQHYHTQWPDGVP---EYTLPLVLTFRKAAAYAKRH 1257
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 108 ITGPLLHVCTAGVGR 122
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 442 AVGPVVVHCSAGVGR 456
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 15
Q9UDR6
ID Q9UDR6 PRELIMINARY; PRT; 1648 AA.
AC Q9UDR6;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE WUGSC:H_DJ1049N15.1 PROTEIN (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99063792; PubMed=9847074;
RA Sulston J.E., Waterston R.;
RT "Toward a complete human genome sequence.";
RL Genome Res. 8:1097-1108(1998).
RN [2]
RP SEQUENCE FROM N.A.
RA Drone K., Le T.;
RT "The sequence of Homo sapiens PAC clone RP5-1049N15.";
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Waterston R.H.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Waterston R.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC006020; AAF03527.1; -.
DR HSP; P18052; IYFO.
DR InterPro; IPR000387; Tyr_phosphatase.
DR InterPro; IPR000242; Tyr_prot_phptase.
DR Pfam; PF00102; Y_phosphatase; 2.
DR PRINTS; PR00700; PRTYPHPTASE.
DR SMART; SM00194; PTPC; 2.
DR PROSITE; PS00383; TYR_PHOSPHATASE_1; 1.
```

```
DR PROSITE; PS00056; TYR_PHOSPHATASE_2; 2.
DR PROSITE; PS00055; TYR_PHOSPHATASE_PTP; 2.
KW Hydrolase.
FT NON_TER .1 1
SQ SEQUENCE 1648 AA; 180082 MW; D4D10F9BDD4A6A55 CRC64;

Query Match 35.4%; Score 238; DB 4; Length 1648;
Best Local Similarity 37.0%; Pred. No. 5e-18;
Matches 50; Conservative 24; Mismatches 43; Indels 18; Gaps 5;

QY 1 DFWGMMNENNCVIAITREIEGGVIKCCSYWPVSLKEPLEFKHFLHLENFQITQYFVI 60
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 143 DFWRMIEHNVEVIVMITNLLEKGRKCDQYWPAGSE--EYGNFLVTQKSVQVLAYTV 1200
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 61 RIFOI----VKKSTGKSHS-----VKHLQFIKWPDHGTPTASVDFFIKYVRYVRKS-----H 107
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1201 RNFTLRNTKIKKSGKGRVVTQHYHTQWPDGVP---EYSLPLVLTFRKAAAYAKRH 1257
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 108 ITGPLLHVCTAGVGR 122
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1258 AVGPVVVHCSAGVGR 1272
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Search completed: March 30, 2002, 08:24:43
Job time: 655 sec
```



**THIS PAGE BLANK (USPTO)**